



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<b>(51) International Patent Classification <sup>6</sup>:</b> <b>C12Q 1/68, C12P 19/34, C07H 21/02, 21/04</b>	<b>A1</b>	<b>(11) International Publication Number:</b> <b>WO 98/07887</b> <b>(43) International Publication Date:</b> 26 February 1998 (26.02.98)
<b>(21) International Application Number:</b> PCT/US97/14892 <b>(22) International Filing Date:</b> 22 August 1997 (22.08.97)  <b>(30) Priority Data:</b> 60/023,438                      23 August 1996 (23.08.96)                      US  <b>(71) Applicant:</b> THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 22nd floor, 300 Lakeside Drive, Oakland, CA 94612-3550 (US).  <b>(72) Inventors:</b> FREIMER, Nelson, B.; 630 29th Street, San Francisco, CA 94131 (US). LEON, Pedro; Centro de Investigaciones Biologia, University of Costa Rica, P.O. Box 2060, San Jose (CR). REUS, Victor, I.; 1214 Third Avenue, San Francisco, CA 94122 (US). SANDKUJIL, Lodewijk, A.; Voorstraat 27A, NL-2611 JK Delft (NL). McINNES, Lynne, Allison; 1599 Shrader Street, San Francisco, CA 94117 (US). SERVICE, Susan, K.; 816 Maher Road, Watsonville, CA 95076 (US).  <b>(74) Agents:</b> HUGHES, Melya, J.; Cooley Godward LLP, 3000 El Camino Real, Five Palo Alto Square, Palo Alto, CA 94306-2155 (US) et al.	<b>(81) Designated States:</b> AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).  <b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
<b>(54) Title:</b> METHODS FOR TREATING BIPOLAR MOOD DISORDER ASSOCIATED WITH MARKERS ON CHROMOSOMES 18p		
<b>(57) Abstract</b>  The present invention is directed to methods of detecting the presence of a bipolar mood disorder susceptibility locus in an individual, comprising analyzing a sample of DNA for the presence of a DNA polymorphism on the short arm of chromosome 18 between the telomere and D18S481, wherein the DNA polymorphism is associated with a form of bipolar mood disorder. The invention for the first time provides strong evidence of a susceptibility gene for bipolar mood disorder that is located in the terminal 5 cM region of the short arm of chromosome 18. The disclosure describes the use of linkage analysis and genetic markers in this 5 cM region to fine map the region and the use of genetic markers to genetically diagnose (genotype) bipolar mood disorder in individuals, to confirm phenotypic diagnoses of bipolar mood disorder, to determine appropriate treatments for patients with particular genotypic subtypes. Isolated polynucleotides useful for genetic linkage analysis of BP-I and methods for obtaining such isolated polynucleotides are also described.		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

1 METHODS FOR TREATING BIPOLAR MOOD DISORDER  
2 ASSOCIATED WITH MARKERS ON CHROMOSOME 18p  
3

4 ACKNOWLEDGEMENTS

5 This invention was made with Government support under Grant Nos. RO1-MH49499,  
6 K21MH00916, awarded by the NIH. The U.S. Government has certain rights in this  
7 invention.  
8

9 INTRODUCTION

10  
11 Background  
12

13 **BIPOLAR MOOD DISORDER (BP)**

14 Manic-depressive illness, or bipolar mood disorder (BP), is characterized by episodes  
15 of elevated mood (mania) and depression and is among the most prevalent and potentially  
16 devastating of psychiatric syndromes. The most severe and clinically distinctive forms of BP  
17 are BP-I (severe bipolar mood disorder) and SAD-M (schizoaffective disorder manic type),  
18 and are characterized by at least one full episode of mania, with or without episodes of major  
19 depression (defined by lowered mood, or depression, with associated disturbances in  
20 rhythmic behaviors such as sleeping, eating, and sexual activity). A milder form of BP is  
21 BP-II, bipolar mood disorder with hypomania and major depression. BP-I often co-  
22 segregates in families with more etiologically heterogeneous syndromes, such as unipolar  
23 major depressive disorder (MDD), which is a more broadly defined phenotype. See  
24 McInnes, L.A. and Freimer, N.B., Mapping genes for psychiatric disorders and behavioral  
25 traits, Curr. Opin. in Genet. and Develop., 5:376-381 (1995).  
26

## 1           TREATMENT OF INDIVIDUALS WITH BIPOLAR MOOD DISORDER

2           An estimated 2-3 million people in the United States are affected by BP-I. Currently,  
3 individuals are typically evaluated for bipolar mood disorder using the clinical criteria set  
4 forth in the most current version of the American Psychiatric Association's Diagnostic and  
5 Statistical Manual of Mental Disorders (DSM). Many drugs have been used to treat  
6 individuals diagnosed with bipolar mood disorder, including lithium salts, carbamazepine and  
7 valproic acid. However, none of the currently available drugs is able to treat every  
8 individual diagnosed with severe BP-I (termed BP-I) and drug treatments are effective in only  
9 approximately 60-70% of individuals diagnosed with BP-I. Moreover, it is currently  
10 impossible to predict which drug treatments will be effective in particular BP-I affected  
11 individuals. Commonly, upon diagnosis affected individuals are prescribed one drug after  
12 another until one is found to be effective. Early prescription of an effective drug treatment is  
13 critical for several reasons, including the avoidance of extremely dangerous manic episodes  
14 and the risk of progressive deterioration if effective treatments are not found. Also,  
15 appropriate treatment may prevent depressive episodes in BP-I individuals; these episodes are  
16 also dangerous and are characterized by a high suicide rate. The high prevalence of the  
17 disorder, together with frequent occurrence of hospitalizations, psychosocial impairment,  
18 suicide and substance abuse, has made BP-I a major public health concern.

## 20           Genetic Basis for Bipolar Mood Disorder

21           Mapping genes for common diseases believed to be caused by multiple genes, such as  
22 BP-I, may be complicated by the typically imprecise definition of phenotypes, by etiologic  
23 heterogeneity and by uncertainty about the mode of genetic transmission of the disease trait.  
24 With psychiatric disorders there is even greater ambiguity in distinguishing between  
25 individuals who likely carry an affected genotype from those who are genetically unaffected.  
26 For example, one can define an affected phenotype for BP by including one or more of the  
27 broad grouping of diagnostic classifications that constitute the mood disorders: BP-I, SAD-  
28 M, MDD, and BP-II.

29           Thus, one of the greatest difficulties facing psychiatric geneticists is uncertainty  
30 regarding the validity of phenotype designations, since clinical diagnoses are based solely on

1 clinical observation and subjective reports. Also, with complex traits such as psychiatric  
2 disorders, it is difficult to map the trait-causing genes genetically because: (1) the BP-I  
3 phenotype doesn't exhibit classic Mendelian recessive or dominant inheritance patterns  
4 attributable to a single genetic locus, (2) there may be incomplete penetrance i.e., individuals  
5 who inherit a predisposing allele may not manifest the disease; (3) the phenocopy  
6 phenomenon may occur, i.e., individuals who do not inherit a predisposing allele may  
7 nevertheless develop the disease due to environmental or random causes; (4) genetic  
8 heterogeneity may exist, in which case mutations in any one of several genes may result in  
9 identical phenotypes.

10 The existence of one or more major genes associated with BP-I and with a clinically  
11 similar diagnostic category, SAD-M (schizoaffective disorder manic subtype), is supported by  
12 segregation analyses and twin studies (Bertelson et al., 1977; Freimer and Reus, 1992; Pauls  
13 et al., 1992). However, efforts to identify the chromosomal location of BP-I genes have  
14 yielded disappointing results in that reports of linkage between BP-I and markers on  
15 chromosomes X and 11 could not be independently replicated nor confirmed in the re-  
16 analyses of the original pedigrees (Baron et al., 1987; Egeland et al., 1987; Kelsoe et al.,  
17 1989; Baron et al., 1993). The possible localization of BP genes on chromosomes 18  
18 (pericentromeric region) and 21q has been suggested, but in both cases the proposed  
19 candidate region is not well defined and there is equivocal support for either location  
20 (Berrettini et al. (1994) Proc. Natl. Acad. Sci. USA, 91, 5918-5921, Murray, J.C., et al.  
21 (1994) Science 265, 2049-2054; Pauls et al., Am. J. Hum. Genet. 57:636-643 (1995); Maier  
22 et al., Psych. Res. 59:7-15 (1995); Straub et al., Nature Genet., 8:291-296 (1994)). Recent  
23 investigations have led to the isolation of chromosome 18-specific brain transcripts which  
24 have been suggested to be positional candidates for bipolar disorder (Yoshikawa et al., Am.  
25 J. Med. Gen. 74, 140-149 (1997)).

26 Despite abundant evidence that BP has a major genetic component, linkage studies  
27 have not yet succeeded in definitively localizing a BP gene. This is mainly because mapping  
28 studies of psychiatric disorders have generally been conducted under a paradigm appropriate  
29 for mapping genes for simple Mendelian disorders, namely, using linkage analysis in the  
30 expectation of finding high lod scores that definitively signpost the location of disease genes.

1 The follow up to early BP linkage studies, however, showed that even extremely high lod  
2 scores at a single location can be false positives. See Egeland, et al., Nature 325:783-787  
3 (1987); Baron et al., Nature 326:289-292 (1987); Kelsoe et al., Nature, 342:238-243 (1989);  
4 and Baron et al., Nature Genet. 3:49-55 (1993). These earlier studies used largely  
5 uninformative markers and did not use stringent criteria for identifying affected individuals.  
6

#### 7 LINKAGE DISEQUILIBRIUM ANALYSIS

8 Linkage disequilibrium (LD) analysis is a powerful tool for mapping disease genes  
9 and may be particularly useful for investigating complex traits. LD mapping is based on the  
10 following expectations: for any two members of a population, it is expected that  
11 recombination events occurring over several generations will have shuffled their genomes, so  
12 that they share little in common with their ancestors. However, if these individuals are  
13 affected with a disease inherited from a common ancestor, the gene responsible for the  
14 disease and the markers that immediately surround it will likely be inherited without change,  
15 or IBD ("identical by descent"), from that ancestor. The size of the regions that remain  
16 shared (i.e. IBD) are inversely proportional to the number of generations separating the  
17 affected individuals and their common ancestor. Thus, "old" populations are suitable for fine  
18 scale mapping and recently founded ones are appropriate for using LD to roughly localize  
19 disease genes more approximately (Houwens et al., 1994, in particular Fig. 3 and  
20 accompanying text). Because isolated populations typically have had a small number of  
21 founders, they are particularly suitable for LD approaches, as indicated by several successful  
22 LD studies conducted in Finland (de la Chapelle, 1993).

23 LD analysis has been used in several positional cloning efforts (Kerem et al., 1989;  
24 MacDonald et al., 1992; Petrukhin et al., 1993; Hastbacka et al., 1992 and 1994), but in  
25 each case the initial localization had been achieved using conventional linkage methods.  
26 Positional cloning is the isolation of a gene solely on the basis of its chromosomal location,  
27 without regard to its biochemical function. Lander and Botstein (1986) proposed that LD  
28 mapping could be used to screen the human genome for disease loci, without conventional  
29 linkage analyses. This approach was not practical until a set of mapped markers covering

1 the genome became available (Weissenbach et al., 1992). The feasibility of genome  
2 screening using LD mapping is now demonstrated by the applicants.

3 Identification of the chromosomal location of a gene responsible for causing severe  
4 bipolar mood disorder can facilitate diagnosis, treatment and genetic counseling of  
5 individuals in affected families.

6 Due to the severity of the disorder and the limitations of a purely phenotypic  
7 diagnosis of BP-I, there is a tremendous need to subtype individuals with BP-I genetically to  
8 confirm clinical diagnoses and to determine appropriate therapies based on their genotypic  
9 subtype.

10

#### 11 SUMMARY OF THE INVENTION

12 The present invention comprises using genetic linkage and haplotype analysis to  
13 identify an individual having a bipolar mood disorder gene on the short arm of chromosome  
14 18. In addition, the present invention provides markers linked to a gene responsible for  
15 susceptibility to bipolar mood disorder that will enable researchers to focus future analysis on  
16 that small chromosomal region and will accelerate the sequencing of a bipolar mood disorder  
17 gene located at 18p.

18 The present invention provides, for the first time, a localization of a BP-I  
19 susceptibility locus to a 300 to 500 kb region of the short arm of chromosome 18.

20 The present invention is directed to methods of detecting the presence of a bipolar  
21 mood disorder susceptibility locus in an individual, comprising analyzing a sample of DNA  
22 for the presence of a DNA polymorphism on the short arm of chromosome 18 between  
23 SAVA5 and ga203, wherein the DNA polymorphism is associated with a form of bipolar  
24 mood disorder. The invention includes the use of genetic markers in the roughly 500 kb  
25 region between the SAVA5 locus and the ga203 locus, inclusive, to diagnose bipolar mood  
26 disorder genetically in individuals and to confirm phenotypic diagnoses of bipolar mood  
27 disorder. Preferably, the sample of DNA is analyzed for the presence of a DNA  
28 polymorphism on the short arm of chromosome 18 in the roughly 300 kb region between  
29 D18S1140 and W3422.

1 In a further embodiment, the invention provides methods of classifying subtypes of  
2 bipolar mood disorder by identifying one of more DNA polymorphisms located within the  
3 500 kb region between SAVA5 and ga203 loci, inclusive, on the short arm of chromosome  
4 18 and analyzing DNA samples from individuals phenotypically diagnosed with bipolar mood  
5 disorder for the presence or absence of one or more of said DNA polymorphisms.  
6 Preferably, the sample of DNA is analyzed for the presence or absence of one or more of  
7 said DNA polymorphisms in the roughly 300 kb region between D18S1140 and W3422 on  
8 the short arm of chromosome 18.

9 In yet a further embodiment, the methods of the invention include a method of  
10 treating an individual diagnosed with bipolar mood disorder comprising identifying one or  
11 more DNA polymorphisms located within the 500 kb region of chromosome 18 between  
12 SAVA5 and ga203, analyzing DNA samples from individuals phenotypically diagnosed with  
13 bipolar mood disorder for the presence or absence of one or more of the DNA  
14 polymorphisms, and selecting a treatment plan that is most effective for individuals having a  
15 particular genotype within the 500 kb region of chromosome 18 between SAVA5 and ga203.  
16 Preferably, the sample of DNA is analyzed for the presence or absence of one or more DNA  
17 polymorphisms in the roughly 300 kb region between D18S1140 and W3422 on the short  
18 arm of chromosome 18.

19

20

#### BRIEF DESCRIPTION OF THE DRAWINGS

21

22

23

24

25

26

27

28

29

30

FIG. 1 is a pedigree chart showing two families, CR001 and CR004. Affected individuals are denoted by black symbols, deceased individuals by a diagonal slash. A schematic of each individual's haplotype (where available) is shown below the ID number. Recombinations are denoted by "-x"; consanguineous marriages by a double bar, and the conserved haplotype as black shading within the haplotype bars. The larger conserved region for CR004 is stippled, the larger conserved region for CR001 is indicated by a dashed outline. An "I" underneath the haplotype bars indicates inferred haplotype. A "?" indicates phase is uncertain. The connection between CR001 and CR004, dating to an 18th Century founding couple, is indicated by the dashed lines joining individuals III-6 and I-4.

1        **FIG. 2** is a table of lod scores for markers covering the entire human genome that  
2 exceeded the arbitrary coverage thresholds. Lod scores are shown for two markers on  
3 chromosome 18: D18S59 and D18S1105.

4  
5        **FIG. 3** depicts the extent of marker coverage used in the pedigree genome screening  
6 study for each chromosome. Coverage is defined as regions for which a lod score of at least  
7 1.6 would have been detected (in the combined data set) for markers truly linked to BP-I  
8 under the model employed. Areas that remain uncovered (at this threshold) are unshaded.  
9 Markers for which lod scores were obtained that exceeded the empirically determined  
10 coverage thresholds in CR001, CR004, or the combined data set, are shown at their  
11 approximate chromosomal location. The symbols to the right of the chromosome indicate the  
12 thresholds exceeded at that marker: a circle signifies that the lod score at a marker exceeded  
13 the threshold of 0.8 in CR001, a diamond signifies that the lod score exceeded the threshold  
14 of 1.2 in CR004, and a star signifies that the lod score exceeded the threshold of 1.6 in the  
15 combined data set.

16  
17        **FIGS. 4A and 4B** depicts the Lod score for the maximum likelihood estimate of theta  
18 in the combined sample for the 473 microsatellite markers typed in the pedigree genome  
19 screen. The MLEs of theta were appointed to the following categories:  $\theta < 0.10$ ;  $0.10$   
20  $\leq \theta \leq 0.40$ ;  $\theta \geq 0.40$ . Note that the scale for the x-axis (distance from pter)  
21 changes with chromosomes.

22  
23        **FIG. 5** is a portion of an integrated map of the 5 cM 18pter region of chromosome  
24 18.

25  
26        **FIGS. 6A, 6B and 6C** are a list of markers on chromosome 18, with map positions  
27 noted.

28  
29        **FIG. 7** describes 18p allele frequencies for disease chromosomes (aff 105) versus  
30 nontransmitted chromosomes (ntrans) and samples from a control population of Costa Rican

1 students and their parents (control). The name of each marker used in this study is indicated  
 2 on the left. The second column of numbers refers to allele length in base pairs.

3  
 4 **FIG. 8** depicts haplotype analysis of individuals affected with BP-I. The column  
 5 labelled 18p refers to the patient identifier, and each patient identifier is repeated with 2 rows  
 6 to indicate allele results with each of the patient's two copies of chromosome 18. The  
 7 columns labelled "PANR" and "MANR" refer to the paternal and maternal identifiers,  
 8 respectively, associated with the particular patient, other than 0, 1 and 2, which indicate that  
 9 parental samples were not available. The column headings to the right of "PANR" and  
 10 "MANR" columns represent names of specific markers in the 18p region that were used in  
 11 the haplotype analysis. The markers are listed in the order they appear on chromosome 18.  
 12 The allele length (in base pairs) is indicated under the column heading each marker for a  
 13 particular patient. In the column to the immediate right of each marker column, a "1"  
 14 indicates that the phase is known, i.e., that it is known whether a particular allele is inherited  
 15 from the paternal or maternal chromosome, and a "0" indicates that the phase is not  
 16 definitely known. The shaded horizontal bars depict haplotypes of at least three markers  
 17 which include a 154 allele length at D18S59, other than patients 218, 225, 232, 234, 311,  
 18 314 and 458, where the stippled region depicts small sections that do not have the 154 allele  
 19 at D18S59. The hatched regions depict uncertainty as to whether the individual has the  
 20 affected haplotype, as the phase is not known with certainty. In addition, the presence of an  
 21 allele length of 232 (or 234) with marker ta201 is thought to result from a highly mutable  
 22 allele and may not be distinct from the 230 allele. Similarly, the 202 allele at ca212 may not  
 23 be distinct from the 200 allele at ca212. Patients 246, 247, 248, 311, 316, 367, 384, 501,  
 24 531, 587, 536, 684, 667 and 669 exhibit a 242, 244, 250, 252 or 214 allele at marker ta201  
 25 which indicates a potential marker location. Patients 488, 435 and 236 exhibit haplotypes  
 26 that are distinct from the pedigrees that were analyzed.

27  
 28 **FIG. 9** depicts haplotype analysis of nontransmitted chromosomes from parents of  
 29 individuals affected with BP-I. The labels "ERSN" and "KID" refer to the parental and  
 30 patient identifiers, respectively. As above, allele length is provided in base pairs below each

1 marker with an indication as to whether phase was known (1) or not known (0) given to the  
2 right of these values. The markers, shading and allele characteristics described for Figure 8  
3 also apply to this figure.

4  
5  
6 **FIG. 10** depicts haplotype analysis of control samples obtained from an unscreened  
7 population of students of the University of Costa Rica and their parents representing the  
8 general population. Identifiers are provided in the column headed "cont", allele length and  
9 phase determination given in the remainder of the table. The markers, shading and allele  
10 characteristics described for Figure 8 also apply to this figure. Complete data for all  
11 markers are not given as indicated by blank boxes, or the terms "miss" or "missing".

12  
13 **FIG. 11** depicts Ancestral Haplotype Reconstruction results in disease chromosomes.

#### 14 15 DESCRIPTION OF SPECIFIC EMBODIMENTS

16 The recent availability of highly polymorphic, genetically mapped markers covering  
17 the human genome (Weissenbach, J., et al. (1996) Nature 359, 794-801, Murray, J.C., et al.  
18 (1994) Science 265, 2049-2054, Gyapay, G., et al. (1994) Nature Genet 7,246-339) has  
19 allowed the development of a multi-stage paradigm for mapping genes for complex traits. In  
20 the first stages, complete genome screening (e.g. through lod score analysis) is used to  
21 identify possible localizations for disease genes. Subsequently, the regions highlighted by the  
22 screening study are more intensively investigated to confirm the initial localizations and  
23 delineate clear candidate regions. Finally, fine mapping methods (such as haplotype or  
24 linkage disequilibrium (LD) analysis) or candidate gene approaches are used for positional  
25 cloning of disease genes.

26 Our genome screening study for BP employed the following strategies. Unlike  
27 previous genetic studies of BP, only those individuals with the most severe and clinically  
28 distinctive forms of BP (BP-I and schizoaffective disorder manic type, SAD-M) were  
29 considered as affected, rather than including those diagnosed with a milder form of BP (BP-  
30 II) or with unipolar major depressive disorder (MDD). Two large pedigrees (CR001 and

1 CR004) were selected from a genetically homogeneous population, that of the Central Valley  
2 of Costa Rica (as described in Escamilla, M.A., et al., (1996) Neuropsychiat. Genet. 67,  
3 244-253, and in Freimer, N.B., et al. (1996) Neuropsychiat. Genet. 67, 254-263, both  
4 incorporated by reference herein). The entire human genome was screened for linkage using  
5 mapped microsatellite markers and a model for genetic analysis in which most of the linkage  
6 information was derived from affected individuals. The goal of this stringent linkage  
7 analysis was to identify all regions potentially harboring major genes for BP-I in the study  
8 population. Empirically determined lod score thresholds (using linkage simulation analyses)  
9 were derived, to suggest regions worthy of further investigation.

10 Identification of all suggestive regions and weighing the relative importance of  
11 findings required complete screening of the genome. The coverage approach was developed  
12 to gauge the progress of this effort. Conventionally, the thoroughness of genome screening  
13 is evaluated by excluding genome regions from linkage under given genetic models. This  
14 approach, which is highly sensitive to misspecification of genetic models, may be poorly  
15 suited for genome screening studies of complex traits; it is tied to the expectation of finding  
16 linkage at a single locus and demonstrating absence of linkage at all other locations in the  
17 genome. Additionally, exclusion analyses do not differentiate between genome regions  
18 where linkage is not excluded because markers are uninformative in the study population  
19 from those in which the genotype data are simply ambiguous. In contrast, the coverage  
20 approach is designed for studies aimed at genome screening rather than for studies where the  
21 goal is to demonstrate a single unequivocal linkage finding, and it provides explicit data  
22 regarding the informativeness of markers in the study pedigrees. Its use lessens the  
23 possibility that one would prematurely dismiss a given genome region as being unpromising  
24 for further study.

25 Because the exact genetic length of chromosomes is not clearly established, it is  
26 impossible to be certain that one has screened the entire genome. Although we report  
27 coverage of about 94% of the genome (under the 90% dominant model) at the thresholds  
28 described above, this probably represents an underestimate. The remaining coverage gaps in  
29 our study occur predominantly at or near telomeres; as the upper bound estimates for the

1 length of each chromosome were used, it is likely that the actual coverage gaps in these  
2 regions are smaller than our conservative assessment.

3 The presence of consistently positive lod scores over a given region was considered to  
4 be of greater significance than isolated peak lod scores. Such clustering suggests true co-  
5 segregation of markers and phenotypes (i.e. alleles are shared identically by descent rather  
6 than identically by state) and is more readily observed in analyses of a few large pedigrees  
7 (as in our study) than in examination of several smaller families. The data presented herein  
8 indicates clustering of positive lod scores in the region of the telomere of 18p.

9 The genome screen was conducted in two stages. The Stage I screen identified areas  
10 suggestive of linkage, so that those areas could be saturated with available markers, and so  
11 that regions, referred to as 'coverage gaps', could be pinpointed where markers were  
12 insufficiently informative in our sample to detect evidence of linkage. The Stage II screen  
13 followed up on regions flanking each marker that yielded peak lod scores approximately  
14 equal to or greater than the thresholds used for the coverage calculations, which were  
15 deemed regions of interest, and filled in coverage gaps. The results of the complete genome  
16 screen (Stages I and II) using 473 markers is described below.

17 In addition, linkage disequilibrium analysis of an independently collected sample of 48  
18 unrelated BP-I patients was initially conducted. These patients were from the same ancestral  
19 population as the patients in the CR001 and CR004 pedigrees. The LD analysis was  
20 conducted with markers on the short arm of chromosome 18 (18p), in a 5 centimorgan (cM)  
21 region ("5 cM 18pter region") extending from the end of the 18p telomere to a distance of 5  
22 cM along the short arm of chromosome 18 (18p). The LD analysis gave evidence of LD in  
23 this region, particularly at marker D18S59 and also at D18S476. LD analysis of further BP-  
24 I patients from the CRCV with markers in this 5 cM 18pter region was conducted to confirm  
25 and fine map a BP-I gene in this region. This approach, using additional BP-I patients from  
26 this CRCV population and additional markers identifies the region of maximum LD and can  
27 precisely localize a BP-I susceptibility gene.

28 Fine mapping of 5 cM 18pter region resulted in the identification of two DNA  
29 markers (D18S1140 and W3422) defining the boundaries of BP-I as approximately 300 kb,  
30 thus allowing a systematic search for the BP-I gene(s).

1 A conservative approach to linkage analysis was used in that almost all of the  
2 information for linkage is derived from individuals with a severe, narrowly defined  
3 phenotype. While this approach made it very unlikely that lod scores greater than  
4 conventional thresholds of statistical significance (e.g.  $\geq 3$ ) would be obtained, it provided  
5 confidence in the robustness of the most suggestive findings.

6 Direct cDNA selection can be used to isolate segments of expressed DNA from the  
7 300 kb region between D18S1140 and W3422 (M. Lovett, J. Kere, L.M. Hinton, *Proc.*  
8 *Natl. Acad. Sci. USA* 88 9628-9632 (1991); Y.-S. Jou *et al.*, *Genomics* 24 410-413 (1994)).  
9 By using bacterial artificial chromosomes (BAC) (e.g., commercially available from  
10 Research Genetics Inc. Huntsville, Alabama), a group of cDNAs can be identified, and  
11 hybridization and PCR-amplification experiments can be used to determine if these cDNA  
12 segments are derived from the 300 kb region.

13 The cDNAs can then be used to determine whether specific sequences are expressed  
14 at lower levels (or not at all) in affected individuals compared to non-carrier individuals.  
15 Measurement of mRNA levels in lymphoblastoid cell lines can be used as an initial screen.  
16 The cell lines are prepared by drawing blood from individuals, transforming the lymphoblasts  
17 with EBV and growing the immortalized cells in culture. Total RNA and DNA are extracted  
18 from the cultured human lymphoblastoid cell lines. Northern blot hybridization is used to  
19 determine reduced levels of a specific sequence compared to levels from an unaffected, non-  
20 carrier individual as a result of mutations in the BP-I gene on the chromosomes from these  
21 affected individuals which results in decreased levels of mature mRNA and play a primary  
22 role in BP-I. Thus, alterations in gene sequences in affected individuals can be determined.

23 The polymerase chain reaction (PCR) is used to amplify the gene and to determine its  
24 sequence from affected individuals. Sequence comparison with unaffected, non-carrier  
25 individuals is carried out to identify polymorphisms in the gene sequence that are responsible  
26 for BP-I.

27 The identification of the biochemical defect that causes BP-I provides a basis for  
28 treatments for this disease. In addition, knowledge that certain mutations in the gene are  
29 responsible for the disease allows mutation detection tests to be used as a definitive diagnosis  
30 for BP-I.

1           Thus, the present invention allows the isolation of a nucleic acid molecule that can be  
2 used in the identification of the presence (or absence) of a mutation in the BP-I gene a human  
3 and thus can be used in the diagnosis of BP-I or in the genetic counseling of individuals, for  
4 example those with a family history of BP-I (although the general population can be screened  
5 as well). In particular, it should be noted that any mutation in the BP-I gene away from the  
6 normal gene sequence is an indication of a potential genetic flaw; even so-called "silent"  
7 mutations that do not encode a different amino acid at the location of the mutation are  
8 potential disease mutations, since such mutations can introduce into (or remove from) the  
9 gene an untranslated genetic signal that interferes with the transcription or translation of the  
10 gene. Thus, advice can be given to a patient concerning the potential for transmission of BP-  
11 I if any mutation is present. While an offspring with the mutation in question may or may  
12 not have symptoms of BP-I, patient care and monitoring can be selected that will be  
13 appropriate for the potential presence of the disease; such additional care and/or monitoring  
14 can be eliminated (along with the concurrent costs) if there are no differences from the  
15 normal gene sequence. As additional information (if any) becomes available (e.g., that a  
16 given silent mutation or conservative replacement mutation does or does not result in BP-I),  
17 the advice given for a particular mutation may change. However, the change in advice given  
18 does not alter the initial determination of the presence or absence of mutations in the gene  
19 causing BP-I.

20           Generally, mutations are identified in the human gene for use in a method of detecting  
21 the presence of a genetic defect that causes or may cause BP-I, or that can or may transmit  
22 BP-I to an offspring of the human. Initially, the practitioner will be looking simply for  
23 differences from the sequence identified as being normal and not associated with disease,  
24 since any deviation from this sequence has the potential of causing disease, which is a  
25 sufficient basis for initial diagnosis, particularly if the different (but still unconfirmed) gene  
26 is found in a person with a family history of BP-I. As specific mutations are identified as  
27 being positively correlated with BP-I (or its absence), practitioners will in some cases focus  
28 on identifying one or more specific mutations of the gene that changes the sequence of a  
29 protein product of the gene or that results in the gene not being transcribed or translated.  
30 However, simple identification of the presence or absence of any mutation in the gene of a

1 patient will continue to be a viable part of genetic analysis for diagnosis, therapy and  
2 counseling.

3       The actual technique used to identify the gene or gene mutant is not itself part of the  
4 practice of the invention. Any of the many techniques to identify gene mutations, whether  
5 now known or later developed, can be used, such as direct sequencing of the gene from  
6 affected individuals, hybridization with specific probes, which includes the technique known  
7 as allele-specific oligonucleotide hybridization, either without amplification or after  
8 amplification of the region being detected, such as by PCR. Other analysis techniques  
9 include single-strand conformation polymorphism (SSCP), restriction fragment length  
10 polymorphism (RFLP), enzymatic mismatch cleavage techniques and transcription/translation  
11 analysis. All of these techniques are described in a number of patents and other publications;  
12 see, for example, "Laboratory Protocols for Mutation Detection" (1996) Oxford University  
13 Press, Editor: Ulf Landegrun.

14       Depending on the patient being tested, different identification techniques can be  
15 selected to achieve particularly advantageous results. For example, for a group of patients  
16 known to be associated with particular mutations of the gene, oligonucleotide ligation assays,  
17 "mini-sequencing" or allele-specific oligonucleotide (ASO) hybridization can be used. For  
18 screening of individuals who are not known to be associated with a particular mutation,  
19 single-strand conformation polymorphism, total sequencing of genetic and/or cDNA and  
20 comparison with standard sequences are preferred.

21       In many identification techniques, some amplification of the host genomic DNA (or of  
22 messenger RNA) will take place to provide for greater sensitivity of analysis. In such cases  
23 it is not necessary to amplify the entire gene, merely the part of the gene or the specific  
24 location within the gene that is being detected. Thus, the method of the invention generally  
25 comprises amplification (such as via PCR) of at least a segment of the gene, with the  
26 segment being selected for the particular analysis being conducted by the diagnostician.

27       The patient on whom diagnosis is being carried out can be an adult, as is usually the  
28 case for genetic counseling, or a newborn, or prenatal diagnosis can be carried out on a  
29 fetus. Blood samples are usually used for genetic analysis of adults or newborns (e.g.,

1 screening of dried blood on filter paper), while samples for prenatal diagnosis are usually  
2 obtained by amniocentesis or chorionic villus biopsy.

3 Prior to the present invention, affected individuals were prescribed one drug after  
4 another until one was found to be effective. As BP-I was diagnosed using clinical criteria,  
5 no correlation between using a particular drug and its efficacy in a given case was observed.  
6 As a result of the present invention, BP-I subtypes can be diagnosed at the molecular level  
7 and effective treatment predicted.

8 For example, lithium salts, carbamazepine and valproic acid have all been prescribed  
9 for BP-I affected individuals with serendipitous results. An individual can now be diagnosed  
10 with bipolar mood disorder by analyzing genetic material from that individual for the  
11 presence or absence of one or more nucleic acid mutations as described above. As a result  
12 of this diagnosis at the molecular level, an effective treatment can be determined by  
13 collecting data to obtain a statistically significant correlation of a particular treatment with  
14 the different subtypes of BP-I. Thus, the practitioner is able to select a specific drug for the  
15 treatment of a particular sub-type of BP-I and does not merely rely on trial and error.

16 Alternatively, the full-length normal genes for BP-I from humans, as well as shorter  
17 genes that produce functional proteins, can be used to correct BP-I in a human patient by  
18 supplying to the human an effective amount of a gene product of the human gene, either by  
19 gene therapy or by *in vitro* production of the protein followed by administration of the  
20 protein. It should be recognized that the various techniques for administering genetic  
21 materials or gene products are well known and are not themselves part of the invention. The  
22 invention merely involves supplying the genetic materials or proteins identified as a result of  
23 the present invention in place of the genetic materials or proteins previously administered.  
24 For example, techniques for transforming cells to produce gene products are described in  
25 U.S. Patent No. 5,283,185 entitled "Method for Delivering Nucleic Acid into Cells," as well  
26 as in numerous scientific articles, such as Felgner et al., "Lipofection: A Highly Efficient,  
27 Lipid-Mediated DNA-Transfection Procedure," *Proc. Natl. Acad. Sci. U.S.A.*, 84, 7413-  
28 7417 (1987); techniques for *in vivo* protein production are described in, for example,  
29 Mueller et al., "Laboratory Methods - Efficient Transfection and Expression of Heterologous  
30 Genes in PC12 Cells," *DNA and Cell Biol.*, 9(3), 221-229 (1990).

1 Administration of proteins and other molecules to overcome a deficiency disease is  
2 well known (e.g., administration of insulin to correct for high blood sugar in diabetes) that  
3 further discussion of this technique is not necessary. Some modification of existing  
4 techniques may be required for particular applications, but those modifications are within the  
5 skill level of the ordinary practitioner using existing knowledge and the guidance provided in  
6 this specification.

7 The invention now being generally described, the following examples are provided for  
8 purposes of illustration only and are not to be considered to limit the invention.

9  
10

#### 11 EXAMPLES

##### 12 PEDIGREES

13 Two independently ascertained Costa Rican pedigrees (CR001 and CR004) were  
14 chosen because they contained a high density of individuals with BP-I and because their  
15 ancestry could be traced to the founding population of the Central Valley of Costa Rica. The  
16 current population of the Central Valley (consisting of about two million people) is  
17 predominantly descended from a small number of Spanish and Amerindian founders in the  
18 16th and 17th centuries (Escamilla, M.A., et al., (1996) Neuropsychiat. Genet. 67, 244-  
19 253). Studies of several inherited diseases have confirmed the genetic isolation of this  
20 population (Leon, P., et al. (1992) Proc. Natl. Acad. Sci. USA. 89, 5181-5184;  
21 Uhrhammer, N., et al. (1992) Am. J. Hum. Genet. 57, 103-111). An extensive description  
22 of pedigrees CR001 and CR004 has been published (Freimer, N.B., et al. (1996)  
23 Neuropsychiat. Genet. 67, 254-263). In the course of the study, two links between these  
24 pedigrees were discovered. However, the families were analyzed separately because these  
25 links were discovered after the simulation analyses were completed and after the genome  
26 screening study had been initiated.

27 All available adult members of these families were interviewed in Spanish using the  
28 Schedule for Affective Disorders and Schizophrenia Lifetime version (SADS-L) (Endicott, J.  
29 et al, (1978) Arch. Gen. Psych. 35, 837-844). Individuals who received a psychiatric  
30 diagnosis were interviewed again in Spanish by a research psychiatrist using the Diagnostic

1 Interview for Genetic Studies (DIGS) (Nurnberger, J.L. et al. (1994) Arch. Gen. Psychiat.  
2 51, 849-859). This recently developed instrument is similar to, but more detailed than  
3 SADS-L. The interviews and medical records were then reviewed by two blinded best  
4 estimators who reached a consensus diagnosis. The diagnostic procedures are described in  
5 detail in Freimer, N.B., et al. (1996) Neuropsychiat. Genet. 67, 254-263 (incorporated by  
6 reference herein).

7

#### 8 UNRELATED CRCV BP-I PATIENT STUDY

9 BP localizations obtained through the CRCV pedigree studies were confirmed by  
10 genotyping an independently collected sample of 48 unrelated BP-I patients from the CRCV.  
11 In this fine mapping LD analysis, 48 unrelated BP-I patients from the CRCV were identified  
12 and genotyped using microsatellite markers spaced at narrow intervals across chromosome  
13 18. As these patients are descended from the same ancestral population as the patients in the  
14 pedigrees previously studied (CR001 and CR004), many of them should share disease  
15 susceptibility alleles inherited identically by descent (IBD) from one or a few common  
16 ancestors, and linkage disequilibrium (LD) should be present at marker loci surrounding the  
17 disease genes.

18 The sample of 48 BP-I patients included 25 women and 23 men who were recruited  
19 from psychiatric hospitals and clinics in the CRCV. These patients were ascertained only on  
20 the basis of diagnosis and CV ancestry, and were not selected on the basis of history of BP  
21 illness in family members. A structured interview of each patient was conducted by a  
22 psychiatrist, and medical and hospital records were collected. Ascertainment and diagnostic  
23 procedures were as described above. However, in order to lessen further the probability of  
24 phenocopies among this unrelated sample, for which we lacked pedigree information, the  
25 affected phenotype was defined even more narrowly than in the pedigree study. Individuals  
26 considered affected in this study had to have suffered at least two disabling episodes of mania  
27 (requiring hospitalization) and a first onset of the illness before age 45.

28 Genealogical research on each of the 48 BP-I patients confirmed that on average, 70%  
29 of their great-grandparents were born in the CRCV. Individuals whose great-grandparents  
30 were born in the CRCV were considered likely to be descended from the original Spanish

1 and Amerindian founders of the CRCV. Genealogical research showed that 2 patients are  
2 first cousins and the remaining 46 have no relationship within the past 4 generations.

#### 4 GENOTYPING PEDIGREE STUDIES

5 Linkage simulations were used to select the most informative individuals from  
6 pedigrees CR001 and CR004 for genotyping studies (Freimer, N.B., et al. (1996)  
7 Neuropsychiat. Genet. 67, 254-263). Under a 90% dominant model, simulation analyses  
8 with these individuals suggested that evidence of linkage would likely be detected (e.g. a  
9 probability of 92% of obtaining lod > 1.0 in the combined data set) using markers with an  
10 average heterozygosity of 0.75 spaced at 10 cM intervals (as discussed in Freimer, N.B., et  
11 al. (1996) Neuropsychiat. Genet. 67, 254-263). For the Stage I screen, the most  
12 polymorphic markers (307 in total) were chosen, placed at approximately 10 cM intervals on  
13 the 1992 Genethon map (Houwen, R., et al. (1992) Nature 359, 794-801). These markers  
14 were then supplemented by a small number of markers from the Cooperative Human Linkage  
15 Center (CHLC) public database. For the Stage II screen, 166 markers were added from  
16 newer Genethon and CHLC maps as they became available (Murray, J.C. et al. (1994)  
17 Science 265, 2049-2054, Gyapay, G., et al. (1994) Nature Genet. 7, 246-339) and from the  
18 public database of the Utah Center for Genome Research, all of which are publicly available.  
19 DNA samples (from individuals in the CEPH families) that were used for size standards for  
20 Genethon and CHLC markers were included in the experiments to permit comparison of  
21 allele sizes between members of the CRCV population and individuals in the CEPH database.  
22 Genotyping procedures were as described previously (DiRienzo, A. et al. (1994) Proc. Natl.  
23 Acad. Sci. USA 91, 3166-3170 (incorporated by reference herein)). Briefly, one of the two  
24 PCR primers was labeled radioactively using a polynucleotide kinase and PCR products were  
25 run on polyacrylamide gels. Autoradiographs were scored independently by two raters.  
26 Data for each marker were entered into the computer database twice and the resultant files  
27 were compared for discrepancies.

28

# GENOTYPING OF UNRELATED BP-I CRCV PATIENTS

Twenty-seven markers were used to genotype all 48 individuals (as well as 53 individuals used to establish genetic phase) at approximately 5 cM intervals along the entire chromosome 18. It was hypothesized that such a screen would permit the evaluation of evidence in the 18pter region and also to investigate other regions on chromosome 18 in which linkage to BP has been suggested by other groups in other populations. For each individual, two-marker haplotypes in each of the 26 inter-marker intervals were investigated. For 38 of the 48 BP-I patients, genotypes of parents or children were available to assist in phase determination. Because of phase ambiguities in the remaining 10 individuals, minimal and maximal two-marker haplotype sharing was evaluated as follows: (1) Minimal: the number of individuals (and chromosomes) who definitely shared a chromosomal segment defined by a particular pair of alleles (phase known chromosomes) and (2) Maximal: the number of individuals (and chromosomes) who could possibly share a chromosomal segment defined by a particular pair of alleles (includes phase unknown chromosomes). The threshold used to identify areas of high IBD sharing of chromosomes in this initial screen was designated as maximal sharing of a two-marker haplotype by 50% or more of the 48 individuals (or 25% or more of the 96 chromosomes).

Arbitrary thresholds were designated to identify possible areas of high IBD sharing among the 48 patients. Eight of the 26 regions passed this screen. Within each of these 3 regions, one to three additional markers were typed to permit detection of LD, if present, over regions of one to two cM.

A total of 42 chromosome 18 markers were used to genotype the study sample: D18S1140, D18S59, D18S476, D18S481, D18S391, D18S452, D18S843, D18S464, D18S1153, D18S378, D18S53, D18S453, D18S40, D18S66, D18S56, D18S57, D18S467, D18S460, D18S450, D18S474, D18S69, D18S64, D18S1134, D18S1147, D18S60, D18S68, D18S55, D18S477, D18S61, D18S488, D18S485, D18S541, D18S870, D18S469, D18S874, D18S380, D18S1121, D18S1009, D18S844, D18S554, D18S461, D18S70 (from pter to qter). Of these 42 markers, four are located within the 5 cM 18pter region extending from the telomere of 18p to marker D18S481 (inclusive), which is approximately 5 cM from the

1 telomere of 18p. This region is referred to as the 5 cM 18pter region. The four markers  
2 tested in the 5 cM 18pter region are: D18S59, D18S1140, D18S476 and D18S481.

3 For each marker the likelihood that a particular allele (or alleles) is over-represented  
4 on disease chromosomes, as compared to non-disease chromosomes was evaluated. The  
5 results of this likelihood test provide a conservative but powerful measure of LD between  
6 two loci.

#### 7 8 **PEDIGREE STATISTICAL ANALYSES**

9 Two-point linkage analyses were performed for all markers. Marker allele  
10 frequencies were estimated from the combined data set with correction for dependency due to  
11 family relationships (Boehnke, M. (1991) Am. J. Hum. Genet. 48, 22-25). The linkage  
12 analyses for Stages I and II included the 65 individuals who were genotyped as well as an  
13 additional 65 individuals who had been diagnostically evaluated but not genotyped. Only  
14 individuals with BP-I were considered affected with the exception of two persons, one in  
15 each family, who carry diagnoses of schizoaffective disorder manic type (SAD-M). The  
16 SAD-M individuals were included as affected because BP-I and SAD-M are often difficult to  
17 distinguish from each other based on their clinical presentation and course of illness  
18 (Goodwin, F.K. et al. (1990) in Manic Depressive Illness (Oxford University Press, New  
19 York), pp. 373-401; Freimer, N.B et al. (1993) in The Molecular and Genetic Basis of  
20 Neurological Disease, pp. 951-965; Freimer, N.B. et al. (1996) Neuropsychiat. Genet. 67,  
21 254-263; and Freimer, N.B. et al (1996) Nature Genetics 12:436-441, all incorporated by  
22 reference herein). In all, 20 individuals were designated as affected within CR004  
23 (Copeman, J.B., et al. (1995) Nature Genet. 9, 80-85 available for genotyping) and  
24 10 individuals from CR001 (Kelsoe, J.R. et al. (1989) Nature 342, 238-243 available for  
25 genotyping). The phenotype for all other individuals was designated as unknown except for  
26 17 individuals who were designated as unaffected because they had been thoroughly clinically  
27 evaluated, showed no evidence of any psychiatric disorder, and were well beyond the age of  
28 risk (50) for BP-I (linkage simulation studies indicated that these unaffected individuals  
29 contributed little information to the linkage analysis).

1           Linkage analyses were performed using a nearly dominant model (assuming  
2 penetrance of 0.81 for heterozygous individuals of 0.9 for homozygotes with the disease  
3 mutation). This model was chosen from five different single-locus models (ranging from  
4 recessive to nearly dominant) due to its consistency with the segregation patterns of BP in the  
5 two pedigrees and because it had demonstrated the greatest power to detect linkage in  
6 simulation studies (Freimer, N.B., et al. (1996) *Neuropsychiat. Genet.* 67, 254-263). Based  
7 on Costa Rican epidemiological surveys Escamilla, M.A., et al., (1996) *Neuropsychiat.*  
8 *Genet.* 67, 244-253, the population prevalence of BP-I was assumed to be 0.015 (and thus  
9 the frequency of the disease allele was assumed to be 0.003)(based on epidemiological  
10 surveys in Costa Rica, Adis, G. (1992) "Disordenes mentales en Costa Rica: Observaciones  
11 Epidemiologicas," (San Jose, Costa Rica: Editorial Nacional de Salud y Seguridad Social)).  
12 The frequency of BP-I in individuals without the disease allele was conservatively set at 0.01  
13 which effectively specified a population phenocopy rate of 0.67 (i.e., an affected individual  
14 in the general population has a 2/3 probability of being a phenocopy). For multiply affected  
15 families, the probability that a gene segregates is highly increased, which implies that  
16 affected individuals in our study pedigree have a lower probability to be phenocopies than  
17 affected individuals in the general population, particularly those with several affected close  
18 relatives (the exact probabilities are dependent on the degree of relationship between patients  
19 and the number of intervening unaffected individuals). These parameters were chosen to  
20 ensure that most of the linkage information derives from affected individuals. The rationale  
21 for selecting these parameters and results of analyses that demonstrate the conservatism of  
22 this model are described by Freimer, N.B., et al. (1996) *Neuropsychiat. Genet.* 67, 254-263.  
23 The LINKAGE package (Lathrop et al., (1984) *Proc. Natl. Acad. Sci. USA* 81, 3443-3446)  
24 was used for lod score analysis and to obtain maximum likelihood estimates of the marker  
25 allele frequencies, taking into account the existing family relationships (see Boehnke, *Am. J.*  
26 *Hum. Gent.* 48, 22-25 (1991)).

27

#### 28           **UNRELATED BP-I CRCV PATIENT STATISTICAL ANALYSES**

29           A likelihood test of disequilibrium (J. Terwilliger, *Am. J. Hum. Genet.* 56, 777  
30 (1995)) was used to estimate a single parameter, lambda, that quantifies the over-

1 representation of marker alleles on disease chromosomes as compared to non-disease  
2 chromosomes. We chose this method of analysis over another commonly used  
3 disequilibrium analysis method, the transmission disequilibrium test (TDT, R. Spielman et  
4 al., Am. J. Hum. Genet. 52, 506 (1993)) because data from all 48 BP-I patients could be  
5 used in the likelihood approach. Effective use of the TDT requires phase-known,  
6 heterozygous parental chromosomes. We do not have parental genotypes for 20 of the 48  
7 BP-I patients. Simulations indicated that with our data, the likelihood test of disequilibrium  
8 would be more powerful than the TDT. Lambda has been shown to be a superior measure  
9 for LD fine mapping, compared to other frequently used measures, because it is directly  
10 related to the recombination fraction between the disease and the marker loci. Non-disease  
11 chromosomes were chosen from the phase-known chromosomes of parents, spouses and  
12 children of affected individuals, if available. Designation of chromosomes of family  
13 members as non-disease in a disorder such as BP-I, which is not fully penetrant, necessitates  
14 specifying a model of disease transmission. The same model of transmission was employed  
15 in this LD likelihood test as was used in the initial genome screen of the pedigrees CR001  
16 and CR002 described herein. One parameter was specified differently from the genome  
17 screen: the phenocopy rate was set to zero in the LD likelihood analysis. A phenocopy rate  
18 was not specified in the transmission model because the effect of phenocopies will be  
19 "absorbed" by the lambda parameter, in that presence of phenocopies in our sample will  
20 serve to erode the association between marker alleles and disease, and hence reduce the  
21 estimate of lambda.

22

### 23 COVERAGE

24 To access coverage for a marker, the number of informative meioses at the estimated  
25 recombination fraction was calculated using the estimate of the variance (the inverse of the  
26 information matrix) (Petrukhin, K.E. et al. (1993) Genomics 15, 76-85). Alternatively,  
27 when the estimated frequency of recombination was close to 0 or 1, Edwards' equation was  
28 applied to calculate the equivalent number of observations (Edwards, J.H. (1971) Ann. Hum.  
29 Genet. 34, 229-250). These meioses represent the amount of linkage information provided  
30 by the marker, given the pedigree structure and the genetic model applied. Linkage to the

1 marker in question was then assumed and the lod score that would be observed as a disease  
2 gene is hypothetically moved in increments away from that marker was calculated. All  
3 regions around a marker that would have generated a lod score that exceeded our thresholds  
4 for possible linkage (0.8 in CR001, 1.2 in CR004, and 1.6 in the combined data) were  
5 considered covered. These lod score thresholds were derived from simulation analyses  
6 showing the expected distribution of lod scores under linkage and non-linkage (Freimer,  
7 N.B., et al. (1996) Neuropsychiat. Genet. 67, 254-263, and approximately represent a result  
8 that is 250 times more likely to occur in linked simulations than in unlinked simulations.  
9 Coverage maps were constructed (FIG. 1) by superimposing the regions covered by each  
10 marker on the genetic map of each chromosome. At the end of the Stage II screen, a total of  
11 473 microsatellite markers had been typed with genome coverage (in the combined data set)  
12 of over 94%. Possible coverage gaps are indicated by unshaded areas and are mainly  
13 concentrated near telomeres. Because the coverage calculations make use of marker  
14 informativeness within the pedigrees, the coverage approach thus permits detection of  
15 instances where markers with expected high heterozygosities are uninformative in our data  
16 set.

17

#### 18 PEDIGREE LINKAGE ANALYSIS RESULTS

19 Of the 473 microsatellites analyzed with two-point linkage tests, 23 markers exceeded  
20 the empirically determined thresholds designated for the coverage calculations (in either  
21 CR001, CR004, or in the combined data set). The location of these markers, the peak lod  
22 scores obtained in each family and in the combined data set, and the maximum likelihood  
23 estimate of the recombination fraction ( $\theta$ ) at which these lod scores were observed are  
24 indicated in Table 1. The approximate chromosomal locations of these markers are also  
25 depicted in FIG. 1. The distribution of lod scores (for the maximum likelihood estimate of  $\theta$   
26 in the combined data set) across the genome is displayed by chromosome in FIG. 2.

27 The threshold was exceeded for pedigree CR001 in two adjacent markers near the 18p  
28 telomere (D18S59 and D18S1105), but CR004 displayed no suggestion of linkage in this  
29 region.

In the genome screen, the highest lod score observed for family CR001 alone was at D18S59 (1.32 at  $\theta=0.0$ ), located near pter. All affected members of CR001 shared alleles at markers in the 18pter region.

#### UNRELATED BP-I CRCV PATIENT STUDY RESULTS

Out of the forty-two markers tested, eight displayed evidence of over-representation of a particular allele on disease chromosomes. Eight of the 42 markers had  $-2 \cdot \ln(\text{likelihood ratio})$  statistics  $> 1.0$ . Three other markers had  $-2 \cdot \ln(\text{likelihood ratio})$  statistics  $> 0$  and  $< 0.62$ . The results are shown in Table I:

Table I

Marker	Allele Size	Frequency on non-disease Chromosomes	Frequency on Disease Chromosomes
D18S59	154	0.121	0.572
D18S476	271	0.470	0.771
D18S467	172	0.384	0.693
D18S61	177	0.074	0.326
D18S485	182	0.237	0.586
D18S870	179	0.405	0.657
D18S469	234	0.128	0.450
D18S1121	168	0.171	0.553

Evidence for association was found at markers located near the telomere of the short arm of chromosome 18. D18S59 displayed the strongest evidence for LD ( $-2 \cdot \ln(\text{likelihood ratio})$  of 8.3,  $p=0.002$ ) of all the chromosome 18 markers tested. An adjacent marker, D18S476 ( $-2 \cdot \ln(\text{likelihood ratio})$  of 1.3), also provided evidence of LD. In our genome screening pedigree study we observed the single highest lod score for pedigree CR001 of any marker in the entire genome at D18S59. Furthermore, the alleles at D18S59 and D18S476

1 that are over-represented among the BP-I patients from the population sample (154 b.p. and  
2 271 b.p. respectively) are observed in all BP-I patients from pedigree CR001.

3 The LD and pedigree findings in the 5 cM 18pter region denote a clearly delineated  
4 region that contains a BP-I susceptibility locus. This region is distinct from other regions on  
5 chromosome 18 that have been suggested as linked to mood disorder phenotypes (more  
6 broadly defined than BP-I). See FIG. 6A, 6B, 6C. In contrast to previous reports by  
7 Berrettini et al. and Stine et al., suggesting possible linkage between mood disorder and  
8 markers in the pericentromeric region of chromosome 18, our results did not show any  
9 evidence for association of BP-I with any pericentromeric markers (D18S378, D18S53,  
10 D18S453 or D18S40).

11

#### 12 IDENTIFICATION OF NEW MARKERS FROM THE 5 cM 18PTER REGION

13 Cloned human genomic DNA covering the target region is assembled. Microsatellite  
14 sequences from these clones are identified. A sufficient area around the repeat to enable  
15 development of a PCR assay for genomic DNA is sequenced, and it is confirmed that the  
16 microsatellite sequence is polymorphic, as several uninformative microsatellites are expected  
17 in any set. Several methods have been routinely used to identify microsatellites from cloned  
18 DNA, and at this time no single one is clearly preferable (Weber, 1990, Hudson et al.,  
19 1992). Most of these require screening an excessive number of small insert clones or  
20 performing extensive subcloning using clones with larger inserts.

21 New strategies have recently been developed which permit the use of the several  
22 different microsatellites to be found within a single large insert clone without requiring  
23 extensive subcloning. A method for direct identification of microsatellites from yeast  
24 artificial chromosomes (YACs) provides several new markers from the target region. This  
25 procedure is based on a subtractive hybridization step that permits separation of the target  
26 DNA from the vector background. This step is useful because the human DNA (the YAC)  
27 constitutes only a small proportion of the total yeast genomic DNA.

28 YAC clones (with inserts averaging about 750 Kb of human genomic DNA) that span  
29 the 5 cM 18pter region have already been identified by the CEPH/G  n  thon consortium  
30 (Cohen et al., 1993) and are publicly available. The markers from YACs that have been

1 mapped to portions of the candidate region that are not well represented by currently  
2 available markers are first isolated. By typing these markers in the families and the "LD"  
3 sample, as described above, it is possible to narrow the candidate region, perhaps to a size of  
4 less than one to two cM, thus permitting limitation of the segment in which more extensive  
5 mapping efforts are applied.

6 Briefly, the microsatellite identification procedure is performed as follows: A  
7 subtractive hybridization is performed using genomic DNA from a target YAC together with  
8 an equivalent amount of a control DNA. This procedure separates the YAC DNA from that  
9 of the yeast vector. Following the subtraction procedure the subtracted YAC DNA is  
10 purified, digested with restriction enzymes and cloned into a plasmid vector (Ostrander et al.,  
11 1992). The cloned products of each YAC are screened using a CA(15) oligonucleotide probe  
12 (i.e. an oligonucleotide having 15 CA repeats). Each positive clone (i.e. those that contain  
13 TG-repeats) is sequenced to identify primers for PCR to genotype the BP-I samples.

14 An alternative approach, based on using a set of degenerate sequencing primers that  
15 anneal directly to the repeat sequence, permitting direct thermal cycle sequencing (Browne &  
16 Litt, 1992), can also be used.

17 Once the candidate region is narrowed to a size of less than about 500 to 1000 Kb, a  
18 contiguous array (contig) of clones with smaller inserts than YACs, mainly P1 clones, is  
19 developed. P1 clones are phage clones specially designed to accommodate inserts of up to  
20 100 Kb (Shepherd et al., 1994).

21

## 22 DEVELOPMENT OF A PHYSICAL MAP OF THE 5 CM 18PTER REGION

23 In parallel with the genetic mapping, a physical map of the 5 cM 18pter region is  
24 developed. The backbone of this effort is the assembly of contigs of large insert clones.  
25 Low resolution contigs for most of the human genome are already available using the YACs  
26 developed by CEPH (Cohen et al., 1993). Although these have been individually verified  
27 and checked for overlap with other YACs, there is a high rate of chimerism in the YACs and  
28 insufficient evidence to definitively confirm the order of the YACs. In addition, because of  
29 their large size these YACs are particularly cumbersome to work with. Nevertheless, they  
30 provide a useful framework to start constructing high resolution contigs.

1        Once a candidate region of less than about five cM is delineated, the studies to  
2        develop a physical map are commenced. Because of the disadvantages of relying solely on  
3        YACs, and because positional cloning is facilitated by the availability of a higher resolution  
4        map, contigs are generated using P1 clones once the candidate region is narrowed to less  
5        than one Mb, by LD mapping in the expanded population sample using the new markers  
6        identified from the YACs.

7        Once a region of 500-1000 Kb or less is defined, physical mapping and cloning are  
8        computed using P1 clones rather than YACs, and P1 contigs over such a region are  
9        constructed. The P1s are used to identify additional markers for the further positional  
10       cloning steps as well as the screening for rearrangements.

11       The starting point of contig construction is the microsatellite sequences and non-  
12       polymorphic STSs that derive from the few YACs that surround the genetically determined  
13       candidate region. These STSs are used to screen the P1 library. The ends of the P1s are  
14       cloned using inverse PCR and used to order the P1s relative to each other. Amplification in  
15       a new P1 will indicate that it overlaps with the previous one. Fluorescent in situ  
16       hybridization (FISH) permits ordering of the majority of the P1s (Pinkel, 1988; Lichter,  
17       1991). The original set of P1s serves as building blocks of the complete contig; each end  
18       clone is used to re-screen the library and in this way P1s are added to the map.

19       From each P1 additional microsatellites are identified as previously described. This  
20       allows further reduction of the candidate region. When the region is narrowed to less than  
21       one Mb in size, positional cloning efforts are initiated.

## 22       **FINE MAPPING OF 5CM 18PTER REGION**

23       In order to delineate further regions of BP-I susceptibility within the 5 cM 18pter  
24       region, additional unrelated BP-I patients from the CRCV as well as other populations can be  
25       diagnosed and genotyped both with the markers described herein as well as additional  
26       markers in the 5 cM 18pter region that are known as well those yet to be identified.  
27       Additional markers are available from the Cooperative Human Linkage Center (CHLC)  
28       public database, from newer Genethon and CHLC maps as they become available (Murray,  
29       J.C. et al. (1994) Science 265, 2049-2054, Gyapay, G., et al. (1994) Nature Genet. 7,246-  
30       339) and from the public database of the Utah Center for Genome Research (all of which are

1 incorporated by reference herein). The web addresses for Genethon and CHLC are:  
2 Genethon ([http://www.genethon.fr/genethon\\_en.html](http://www.genethon.fr/genethon_en.html)), CHLC  
3 (<http://gopher.chlc.org/HomePage.html>). These databases are all linked, and one of ordinary  
4 skill in the art can readily access the information available from these databases.

5 The markers shown in FIG. 6A, from number 1 to 22 or 23 can be used to genotype  
6 the CRCV pedigrees and unrelated BP-I patients described herein as well as other BP-I  
7 affected individuals and pedigrees. See FIG. 6A (portion of a chromosome 18 map available  
8 from the Whitehead Institute, web address: [http://133.30.8.1:8080/=@@=:www-](http://133.30.8.1:8080/=@@=:www-genome.wi.mit.edu)  
9 [genome.wi.mit.edu](http://133.30.8.1:8080/=@@=:www-genome.wi.mit.edu). (incorporated herein by reference)). The fine mapping techniques  
10 described herein in conjunction with the teachings regarding the 5 cM 18pter region can be  
11 used to narrow the BP-I susceptibility region further.

12 The following markers (listed in order of occurrence from the telomere towards the  
13 centromere) were used to delineate regions of BP-I susceptibility within the 5 cM 18pter  
14 region: SAVA5, ca211, ca212, D18S1140, D18S59, ca231, ta201, AT201, ca225, w3442,  
15 ca213, ga201, ga203, ca219, D18S1105, ca209, ca202, D18S1146, GATA (referred to in the  
16 figures as 166d05) and D18S476. The markers SAVA5, D18S1140, D18S59, ta201, at201,  
17 w3442, ga201, ga203, D18S1105, D18S1146, GATA and D18S476 were used in both the  
18 haplotype analysis (Figure 8) and the AHR analysis (Figure 11) to delineate the BP-I  
19 susceptibility locus to the 500 kb region defined by the markers SAVA5 and ga203 and the  
20 300 kb region defined by D18S1140 and W3422. The other markers were used in both  
21 haplotype and the AHR analyses as confirmatory evidence for the localizations. Blood  
22 samples from 105 affected individuals were tested for the presence of marker haplotypes and  
23 compared to marker haplotypes detected on the non-transmitted chromosome in samples  
24 obtained from the parent(s) of the affected individuals when available (71 cases) or to  
25 markers detected in samples obtained from a control population of students attending the  
26 University of Costa Rica (52 samples). The non-transmitted chromosomes are well matched  
27 as controls allowing the affected haplotype of the transmitted chromosome to be more easily  
28 discerned than through comparison with data obtained from the general population that may  
29 contain individuals who carry the affected haplotype but do not exhibit clinical symptoms of  
30 bipolar mood disorder.

1           Figure 7 provides 18p allele frequencies for disease (aff 105) versus nontransmitted  
2 (ntrans) chromosomes and samples from the control population of students (control). The  
3 name of each marker used in this study is indicated on the left. The second column of  
4 numbers refers to allele length in basepairs. This data provides evidence of over-  
5 representation of a particular allele on disease chromosomes.

6           Figure 8 summarizes the results obtained with affected individuals. The column  
7 labelled 18p refers to the patient identifier, and each patient identifier is repeated to indicate  
8 results with both copies of chromosome 18. The labels "PANR" and "MANR" refer to the  
9 paternal and maternal identifier, respectively, associated with the particular patient, other  
10 than 0, 1 and 2, which indicate that parental samples were not available. The allele length  
11 (base pairs) is indicated under each marker for a particular patient; the length of the  
12 horizontal bar in the figure reflects whether haplotypes are IBD or IBS, with IBD haplotypes  
13 with common ancestors having longer bars than randomly inherited IBS haplotypes. To the  
14 right of each marker, a "1" indicates that the phase is known, i.e., that it is known whether a  
15 particular allele is inherited from the paternal or maternal chromosome, and a "0" indicates  
16 that the phase is not known for sure. The determination of phase allows the practitioner to  
17 conclude that marker alleles are linked in a haplotype on the same disease causing  
18 chromosome.

19           Figure 9 provides similar data for non-transmitted chromosomes obtained from  
20 parental samples. Some individuals exhibited the affected haplotype indicating that the parent  
21 was homozygous; however, these regions of identity were typically much shorter than those  
22 regions observed in affected individuals, indicating that they were IBS.

23           Figure 10 similarly provides data for an unscreened population of students  
24 from the University of Costa Rica and their parents (52 samples in total). The data  
25 demonstrate that there is a lower incidence of the affected haplotype in the general population  
26 as compared with Figure 8 and that the affected haplotype is typically shorter similar to the  
27 results obtained with non-transmitted chromosomes. However, the results for the general  
28 population is less distinctive than that observed for non-transmitted chromosomes in allowing  
29 one to map the affected haplotype.

1           Comparison of the affected haplotype with non-transmitted chromosome markers  
2 indicate that the region of maximal sharing between affected individuals occurs between  
3 1140t and w3442 on chromosome 18. This region encompasses approximately 300 kb.  
4           The data was analyzed further using Ancestral Haplotype Reconstruction (AHR), a  
5 likelihood method for measuring LD. Data from affected individuals are examined in 2-  
6 marker segments. Within each segment, the multinomial likelihood of each of the possible  
7 ancestral haplotypes giving rise to the observed sample of disease haplotypes is calculated.  
8 This likelihood is calculated assuming some fraction,  $\alpha$ , of disease chromosomes are  
9 associated with this 2-marker segment, and  $(1-\alpha)$  are linked to this segment. These  
10 haplotype likelihoods are weighted by the probability of observing that haplotype in the  
11 population, and summed to create an overall likelihood for the 2-marker segment. This  
12 segment likelihood is compared to the null likelihood, which assumes the disease and  
13 markers are unlinked (and therefore  $\alpha=0$ ), and a LOD score is generated. The LOD score  
14 is maximized over the parameter  $\alpha$ . Details of these calculations are presented in Appendix  
15 A. The results of this analysis are shown in Figure 11. The percentages given above the  
16 diagonal line demarcated by the filled boxes indicate the percentage of disease chromosomes  
17 hypothesized to be true chromosomes from a common founder. For example, 17% of  
18 chromosomes obtained from affected individuals have the 18S59 to W3442 region; i.e., as  
19 each individual has two chromosome copies, 34% of individuals have this region. The  
20 number above each percentage indicates the LOD score. The numbers given below the  
21 diagonal line demarcated by the filled boxes indicate the alleles inherited from a common  
22 founder, with the number prior to the dash indicating the allele of the marker identified in  
23 the column heading and the number following the dash indicating the allele of the marker  
24 identified in the row heading. The marker alleles are referred to as follows:  
25

1	MARKER	#	ALLELE LENGTH
2	SAVA5	2	229
3	CA211	3	195
4	18S1140	2	268
5	18S59	4	154
6	18S59	6	158
7	TA201	2	220
8	TA201	3	230
9	CA231	2	186
10	CA231	4	202
11	AT201	1	170
12	AT201	2	178
13	CA225	1	160
14	CA225	3	172
15	W3442	1	10

16 Blank boxes indicate no positive evidence for linking the indicated region to the affected  
17 chromosome.

18

#### 19 USE OF P1 CLONES TO IDENTIFY CANDIDATE cDNAs FOR SCREENING FOR MUTATIONS 20 IN THE DNA OF BP-I PATIENTS

21

22 The P1 clones described above are used to identify candidate cDNAs. The candidate  
23 cDNAs are subsequently screened for mutations in DNA from BP-I patients. From the  
24 minimal candidate region defined by genetic mapping experiments a segment is left that is  
25 sufficiently large to contain multiple different genes.

26

#### 27 IDENTIFICATION OF CODING SEQUENCES

28 Coding sequences from the surrounding DNA are identified, and these sequences are  
29 screened until a probable candidate cDNA is found. Much of the human genome will be  
30 sequenced over the next few years, in which case it may become feasible to identify coding  
31 sequences through database screening. Candidates may also be identified by scanning

1 databases consisting of partially sequenced cDNAs (Adams et al., 1991), known as expressed  
2 sequence tags, or ESTs. These resources are already largely developed, and include upwards  
3 of 100,000 cDNAs, the majority expressed primarily in the brain. It is not yet clear,  
4 however, that the complete set of cDNAs will be mapped to specific chromosomal locations  
5 in the near future, and that their data will soon be made publicly available. The database can  
6 be used to identify all cDNAs that map to the minimal candidate region for BP-I. These  
7 cDNAs are then used as probes to hybridize to the P1 contig, and new microsatellites are  
8 isolated, which are used to genotype the "LD" sample. Maximal linkage disequilibrium in  
9 the vicinity of one or two cDNAs is identified. These cDNAs are the first ones used to  
10 screen patient DNA for mutations. Database screening has already been used to identify a  
11 gene responsible for familial colon cancer (Papadopolous et al., 1993).

12 Coding sequences are also identified by exon amplification (Duyk et al., 1990;  
13 Buckler et al., 1991). Exon amplification targets exons in genomic DNA by identifying the  
14 consensus splice sequences that flank exon-intron boundaries. Briefly, exons are trapped in  
15 the process of cloning genomic DNA (e.g. from P1s) into an expression vector (Zhang et al.,  
16 1994). These clones are transfected into COS cells, RT-PCR is performed on total or  
17 cytoplasmic RNA isolated from the COS cells using primers that are complementary to the  
18 splicing vector. Exon amplification is tedious but routine; for example, the system developed  
19 by Buckler et al. (1991). This method is probably preferable to another widely used  
20 approach, direct selection, which involves screening cDNAs using large insert clone contigs,  
21 with several steps to maximize the efficiency of hybridization and recovery of the appropriate  
22 hybrid (Lovett et al., 1991). Although direct selection is more efficient than exon  
23 amplification (Del Mastro et al., 1994), it may not be practical as it depends on the candidate  
24 cDNA being expressed in the tissue from which the cDNA library was made; there is no  
25 prior information to indicate the tissue or developmental stage in which BP-I genes would be  
26 expressed.

27 Once cDNAs are identified the most plausible candidates are screened by direct  
28 sequencing, SSCP or using chemical cleavage assays (Cotton et al. 1988).

29 The data are also evaluated for clues to the possible identity or mode of action of BP-  
30 I mutations. For example, it is known that trinucleotide repeat expansion is associated with

1 the phenomenon of anticipation, or the tendency for a phenotype to become more severe and  
2 display an earlier age of onset in the lower generations of a pedigree (Ballabio, 1993).  
3 Several investigators have suggested that segregation patterns of BP-I are consistent with  
4 anticipation (McInnis et al., 1993; Nylander et al., 1994). The apparent transmission of BP-  
5 I, in association with the conserved 18q23 haplotype is constant with anticipation.  
6 Therefore, once the candidate region is narrowed to its minimal extent, the P1 clones are  
7 screened using trinucleotide repeat oligonucleotides (Hummerich et al., 1994). A PCR assay  
8 is developed and patient DNAs are screened for expanded alleles.

9 Genetic and physical data help to map the bipolar mood disorder gene to the 5 cM  
10 18pter region of chromosome 18. New markers from this region are tested in order to locate  
11 the bipolar mood disorder gene in a region small enough to provide higher quality genetic  
12 tests for bipolar mood disorder, and to specifically find the mutated gene. Narrowing down  
13 the region in which the gene is located will lead to sequencing of the bipolar mood disorder  
14 gene as well as cloning thereof. Further genetic analysis employing, for example, new  
15 polymorphisms flanking D18S59 and D18S476 as well as the use of cosmids, yeast artificial  
16 chromosome (YAC) clones, or mixtures thereof, are employed in the narrowing down  
17 process. The next step in narrowing down the candidate region includes cloning of the  
18 chromosomal region 18pter including proximal and distal markers in a contig formed by  
19 overlapping cosmids and YACS. Subsequent subcloning in cosmids, plasmids or phages will  
20 generate additional probes for more detailed mapping.

21 The next step of cloning the gene involves exon trapping, screening of cDNA  
22 libraries, Northern blots or rt PCR (reverse transcriptase PCR) of samples from affected and  
23 unaffected individuals, direct sequencing of exons or testing exons by SSCP (single strand  
24 conformation polymorphism), RNase protection or chemical cleavage.

25 Flanking markers on both sides of the bipolar mood disorder gene combined with  
26 D18S59 and D18S476 or a number of well-positioned markers that cover the chromosomal  
27 region (5 cM 18pter) carrying the disease gene, can give a high probability of affected or  
28 non-affected chromosomes in the range of 80-90% accuracy, depending on the  
29 informativeness of the markers used and their distance from the disease gene. Using current  
30 markers linked to bipolar mood disorder, and assuming closer flanking markers will be

1 identified, a genetic test for families with bipolar mood disorder will be for diagnosis in  
2 conjunction with clinical evaluation, screening of risk and carrier testing in healthy siblings.  
3 In the future, subsequent delineation of closely linked markers which may show strong  
4 disequilibrium with the disorder, or identification of the defective gene, could allow  
5 screening of the entire at-risk population to identify carriers, and provide improved  
6 treatments.

7

#### 8 TREATMENT OF BP-I PATIENTS USING GENOTYPE DATA

9 Using the fine mapping techniques described herein, BP-I susceptibility loci or genes  
10 in the 5 cM 18pter region in particular in the region #1 between SAVA5 and ga203, are  
11 identified and used to genotype patients diagnosed phenotypically with BP-I. Preferably,  
12 markers in the roughly 500 kb region defined by SAVA5 and ga203, inclusive, are used.  
13 More preferably, markers in either the region defined by D18S59 and w3422, inclusive, are  
14 used.

15 Genotyping with the markers described herein as well as additional markers permits  
16 confirmation of phenotypic BP-I diagnoses or assist with ambiguous clinical phenotypes  
17 which make it difficult to distinguish between BP-I and other possible psychiatric illnesses.  
18 A patient's genotype in the 5 cM 18pter region is determined and compared with previously  
19 determined genotypes of other individuals previously diagnosed with BP-I. Once an  
20 individual is genotyped as having a BP-I susceptibility locus in the 5 cM 18pter region, the  
21 individual is treated with any of the known methods effective in treating at least certain  
22 individuals affected with BP-I, such as the administration of lithium salts, carbamazepine or  
23 valproic acid.

24 Studies are conducted correlating effective treatments with BP-I genotypes in the 5  
25 cM 18pter region to determine the most effective treatments for particular genotypes. BP-I  
26 patients can then be genotyped in the 5 cM 18pter region and the statistically most effective  
27 treatment can be determined as a first course of therapy.

28 All publications and patent applications mentioned in this specification are herein  
29 incorporated by reference to the same extent as if each individual publication or patent  
30 application was specifically and individually indicated to be incorporated by reference.

- 1           The invention now being fully described, it will be apparent to one of ordinary skill
- 2   in the art that many changes and modifications can be made thereto without departing from
- 3   the spirit or scope of the appended claims.

referred from file

### Appendix A

Consider the original mutation to have occurred on a chromosomal segment between two markers A and B. At the time the mutation was introduced, there were  $n_a$  alleles at marker A and  $n_b$  alleles at marker B. On the chromosome containing the disease mutation both marker A and marker B carried allele X. The probability that after  $g$  generations an affected individual carrying the original disease mutation would still have allele X at markers A and B is:

$$(1-\theta_1)^g(1-\theta_2)^g + (1-\theta_1)^g(1-(1-\theta_2)^g)f(X_B) + (1-(1-\theta_1)^g)(1-\theta_2)^gf(X_A) +$$

eq (1)  $(1-(1-\theta_1)^g)(1-(1-\theta_2)^g)f(X_A)f(X_B)$

where  $\theta_1$  is the recombination fraction between disease and marker A,  $\theta_2$  is the recombination fraction between disease and marker B,  $g$  is the number of generations since founding (i.e. since the mutation was introduced into the population),  $f(X_A)$  is the population frequency of the X-allele at marker A and  $f(X_B)$  is the population frequency of the X-allele at marker B. This equation includes terms for the possibility of recombination between the markers and the disease locus, with the X-allele at the markers then being identical by state (IBS) rather than IBD. The probabilities of an affected individual with the original mutation having other haplotypes can be formulated similarly. The probability of having allele Z at marker B (where Z is any allele at marker B besides X) would be:

$$(1-\theta_1)^g(1-(1-\theta_2)^g)f(Z_B) + (1-(1-\theta_1)^g)(1-(1-\theta_2)^g)f(X_A)f(Z_B)$$

eq (2)

where  $f(Z_B)$  is the frequency of allele Z at marker B in the population. The probability of having allele Z at marker A (where Z is any allele at marker B besides X) would be :

$$(1-\theta_2)^g(1-(1-\theta_1)^g)f(Z_A) + (1-(1-\theta_1)^g)(1-(1-\theta_2)^g)f(X_B)f(Z_A)$$

eq (3)

where  $f(Z_A)$  is the frequency of allele Z at marker A in the population. Finally, the probability of having allele Z at both markers A and B would be:

$$(1-(1-\theta_1)^g)(1-(1-\theta_2)^g)f(Z_A)f(Z_B)$$

eq (4)

These probabilities assume (1) no interference in recombination and (2) the same marker alleles are present now as were present  $g$  generations ago, in similar frequencies. If, for example, marker A has  $n_a$  alleles and marker B has  $n_b$  alleles, then these probabilities form a  $(n_a) \times (n_b)$  by  $(n_a) \times (n_b)$  transition matrix, with row  $i$  containing the probabilities that founder haplotype  $i$  gave rise to each of the  $(n_a) \times (n_b)$  different haplotypes in  $g$  generations. The rows of this transition matrix sum to 1.

In simulations, the haplotype frequencies in the disease population were formulated using these transition probabilities, assuming the disease arose on a haplotype with the "1" allele at each of the two markers.

Once these transition probabilities are estimated, the likelihood of a particular founder chromosome giving rise to the observed sample of disease haplotypes in  $g$  generations is easily estimated. For example, if one assumed that the disease mutation arose on a chromosome with the X-allele at both markers, the likelihood ( $L_{X-X}$ ) that this chromosome was the founder of the present-day sampled disease chromosomes is given by the multinomial:

$$L_{X-X} = \prod_{i=1}^K (p_{X-X,i})^{Y_i}$$

eq (5)

where  $i$  indexes the  $K$  potential haplotypes for the two markers ( $K = (n_a)(n_b)$ ),  $p_{X-X,i}$  is the probability that the ancestral disease chromosome with the X-allele at both markers gave rise to a haplotype of type  $i$  in  $g$  generations (taken from the transition matrix), and  $Y_i$  is the observed number of haplotype  $i$  in the sample ( $\sum_i(Y_i) = \text{the number of chromosomes in the sample to be analyzed}$ ). The likelihood in eq (5) assumes that all affected individuals are independent. While, after many generations of separation from a common ancestor one might consider these

individuals to be independent, they are in fact related through a complex and unknown pedigree. The simplification of considering individuals as independent makes the likelihood much more tractable to compute.

The  $K$  likelihoods are then summed, and weighted by the probability of observing that particular haplotype in the population to produce an overall likelihood for the 2-marker segment:

$$L = \sum_{i=1}^K f_i L_i$$

eq (6)

where  $f_i$  is the frequency of haplotype  $i$  in the population. This overall likelihood calculation parallels the approach taken by Terwilliger (1995, eq (2)). The haplotype frequencies are estimated from the sample of normal chromosomes. In the event that the haplotype resulting in the largest contribution to the overall likelihood in eq (6) is not observed in the normal sample, the upper 95% confidence interval for this frequency is used, and the remaining haplotype frequencies rescaled accordingly.

This overall likelihood is compared to the null likelihood, which is generated in exactly the same manner, except that it is assumed the markers were unlinked to

the disease locus ( $\theta_1=\theta_2=0.5$  in, for example, eqs (1-4)). The  $\log_{10}$  of this likelihood ratio is a LOD score. One might consider to use in the null likelihood transition probabilities calculated under the assumption of linkage equilibrium. Under this null the cells of the transition matrix are computed by multiplication of allele frequencies, assuming independence of marker loci. These two forms of the null likelihood are equivalent in value for  $g$  of approximately 20 or greater, and for  $g < 20$  the values are nearly equivalent.

Because  $\theta_1$  and  $\theta_2$  are obviously unknown, the putative disease locus is set to be in the middle of the segment and therefore  $\theta_1$  and  $\theta_2$  are one-half the genetic distance (converted to recombination fraction by the Haldane mapping function, (Ott 1991)) between the two marker loci forming the segment. In fact, one could estimate  $\theta_1$  and  $\theta_2$ , or their ratio, and the method could easily be modified to do so, however for our purposes finding a linked segment is suitable.

This basic procedure has been modified to deal with heterogeneity in the sample of disease chromosomes. Not all chromosomes in the disease sample may be true disease chromosomes from a common founder. Individuals heterozygous for the disease mutation will add one chromosome to the disease sample that will not be a true disease chromosome. Additionally, affected individuals not linked to the

particular chromosomal segment being analyzed (either because they are phenocopies or because of locus heterogeneity) will contribute two chromosomes to the affected sample that do not harbor this disease locus. When the null hypothesis of no linkage is not true, some fraction,  $\alpha$ , of the chromosomes in the disease sample will be associated with this chromosomal segment, and  $(1-\alpha)$  will not be associated. We decided to examine  $\alpha$  in steps of 0.1, from 1.0 to 0.0, and for each step in  $\alpha$  produce a new transition matrix under the alternative hypothesis and calculate a LOD score. If we call the transition matrix calculated under the alternative hypothesis (where the disease locus is hypothesized to be in the middle of the 2-marker segment)  $T_a$  and call the transition matrix calculated under the null hypothesis (where the disease locus is unlinked to the 2-marker segment)  $T_n$ , then a new transition matrix for the alternative hypothesis is calculated as:

$$T^* = \alpha T_a + (1 - \alpha) T_n$$

eq (7)

The transition matrix under the null uses  $\alpha=0$ . The LOD score is then maximized over the one parameter  $\alpha$ .

1 WHAT IS CLAIMED IS:

2

3 1. A method of detecting the presence of a bipolar mood disorder susceptibility locus in  
4 an individual comprising:

5 analyzing a sample of DNA from said individual for the presence of a DNA  
6 polymorphism on the short arm of chromosome 18 between SAVA5 and ga203, wherein said  
7 DNA polymorphism is associated with a form of bipolar mood disorder.

8

9 2. The method of claim 1, wherein said DNA polymorphism is located on the short arm  
10 of chromosome 18 between D18S1140 and ga203, inclusive.

11

12 3. The method of claim 1, wherein said DNA polymorphism is located on the short arm  
13 of chromosome 18 between SAVA5 and W3422, inclusive.

14

15 4. The method of claim 1, wherein said DNA polymorphism is located on the short arm  
16 of chromosome 18 between D18S1140 and W3422, inclusive.

17

18 5. The method of claim 1, wherein said DNA polymorphism is located on the short arm  
19 of chromosome 18 between D18S1140 and at201, inclusive.

20

21 6. The method of claim 1, wherein said DNA polymorphism is located on the short arm  
22 of chromosome 18 between D18S1140 and ta201, inclusive.

23

24 7. The method of claim 1, wherein said DNA polymorphism is located on the short arm  
25 of chromosome 18 between D18S59 and ta201, inclusive.

26

- 1 8. The method of claim 1, wherein said analyzing further comprises:  
2 a. obtaining DNA samples from family members of said individual,  
3 b. analyzing said DNA samples from family members for the presence of said DNA  
4 polymorphism, and  
5 c. correlating the presence or absence of the DNA polymorphism with a  
6 phenotypic diagnosis of bipolar mood disorder for said individual and for said family  
7 members.  
8
- 9 9. A method for detecting the presence of a DNA polymorphism linked to a gene  
10 associated with bipolar mood disorder in an individual comprising:  
11 a. typing blood relatives of said individual for a DNA polymorphism located  
12 within a 500kb region of chromosome 18, wherein said region is located between SAVA5  
13 and ga203, inclusive,  
14 b. analyzing a DNA sample from said individual for the presence of said DNA  
15 polymorphism.  
16
- 17 10. A method of genetically diagnosing bipolar mood disorder in an individual  
18 comprising:  
19 a. obtaining a DNA sample from said individual,  
20 b. analyzing said DNA sample for the presence of a DNA polymorphism  
21 associated with bipolar mood disorder, wherein said DNA polymorphism is located within a  
22 500 kb region of chromosome 18, wherein said region is located between SAVA5 and ga203,  
23 inclusive.  
24
- 25 11. A method of confirming a phenotypic diagnosis of bipolar mood disorder in an  
26 individual comprising:  
27 a. obtaining a DNA sample from said individual,  
28 b. analyzing said DNA sample for the presence of a DNA polymorphism  
29 associated with bipolar mood disorder, wherein said DNA polymorphism is located within a

- 1 500 kb region of chromosome 18, wherein said region is located between SAVAS and ga203,  
2 inclusive.  
3
- 4 12. The method of claim 10, wherein said individual has Spanish or Amerindian ancestry.  
5
- 6 13. A method of classifying subtypes of bipolar mood disorder comprising:  
7 a. identifying one or more DNA polymorphisms located within a 500 kb region  
8 of chromosome 18, wherein said region is located between SAVAS and ga203, inclusive; and  
9  
10 b. analyzing DNA samples from individuals phenotypically diagnosed with  
11 bipolar mood disorder for the presence or absence of one of more of said DNA  
12 polymorphisms.  
13
- 14 14. A method of treating an individual diagnosed with bipolar mood disorder comprising:  
15 a. identifying one or more DNA polymorphisms located within a 500 kb region  
16 of chromosome 18, wherein said region is located between SAVAS and ga203, inclusive; and  
17  
18 b. analyzing DNA samples from individuals phenotypically diagnosed with  
19 bipolar mood disorder for the presence or absence of one of more of said DNA  
20 polymorphisms, and  
21 c. selecting a treatment plan that is most effective for individuals having a  
22 particular genotype within said 500 kb region of chromosome 18.  
23
- 24 15. An isolated polynucleotide capable of selectively hybridizing with a DNA sample  
25 from an individual phenotypically diagnosed with severe bipolar mood disorder, wherein said  
26 polynucleotide does not selectively hybridize with a DNA sample from an individual not  
27 affected by severe bipolar mood disorder, wherein said isolated polynucleotide selectively  
28 hybridizes with a complementary polynucleotide within a 500 kb region of chromosome 18,  
29 wherein said region is located between SAVAS and ga203, inclusive.  
30

- 1 16. The isolated polynucleotide of claim 15, wherein said complementary polynucleotide  
2 is within a 500 kb region of chromosome 18, between SAVA5 and ga203, inclusive.  
3

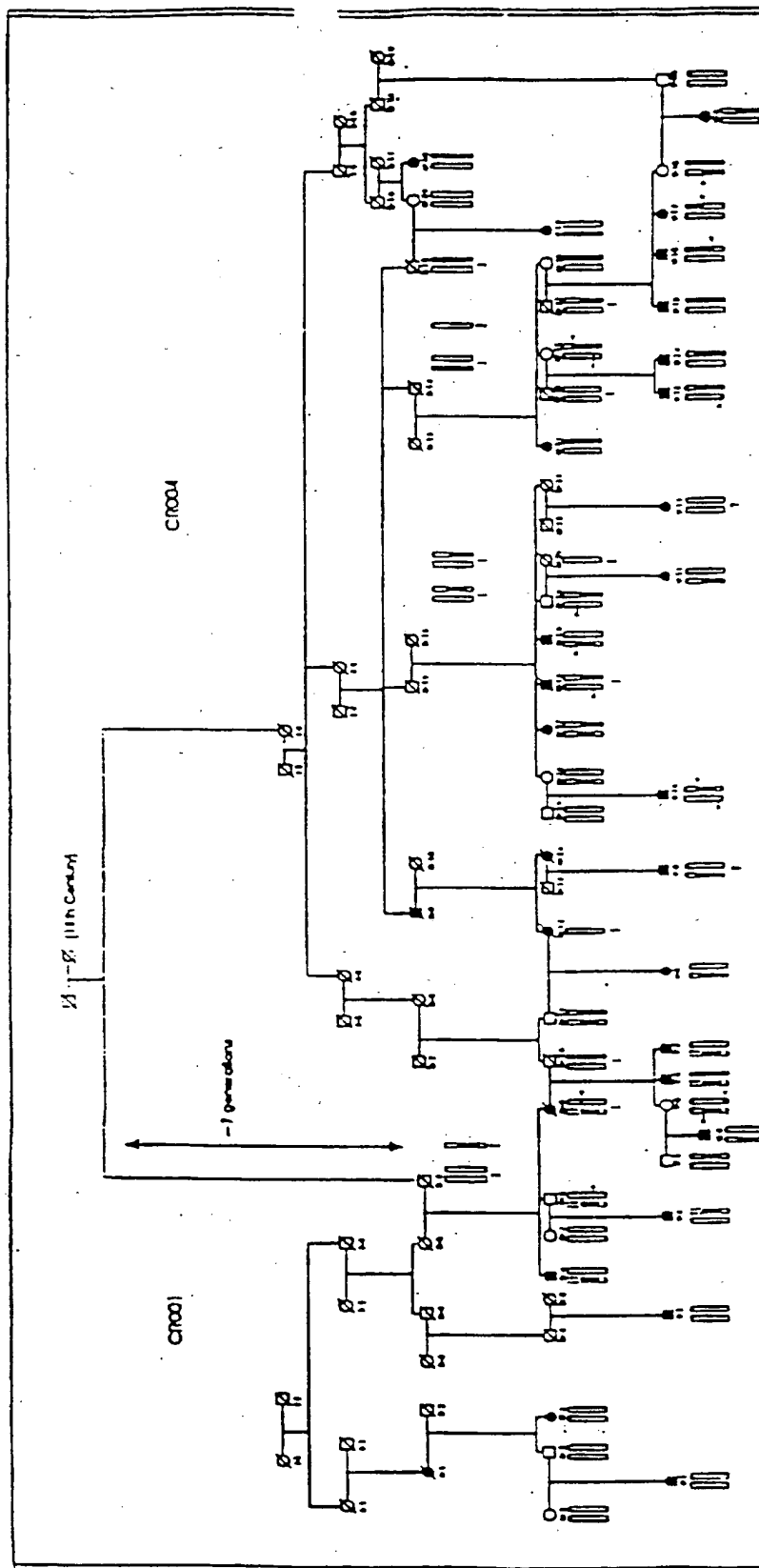


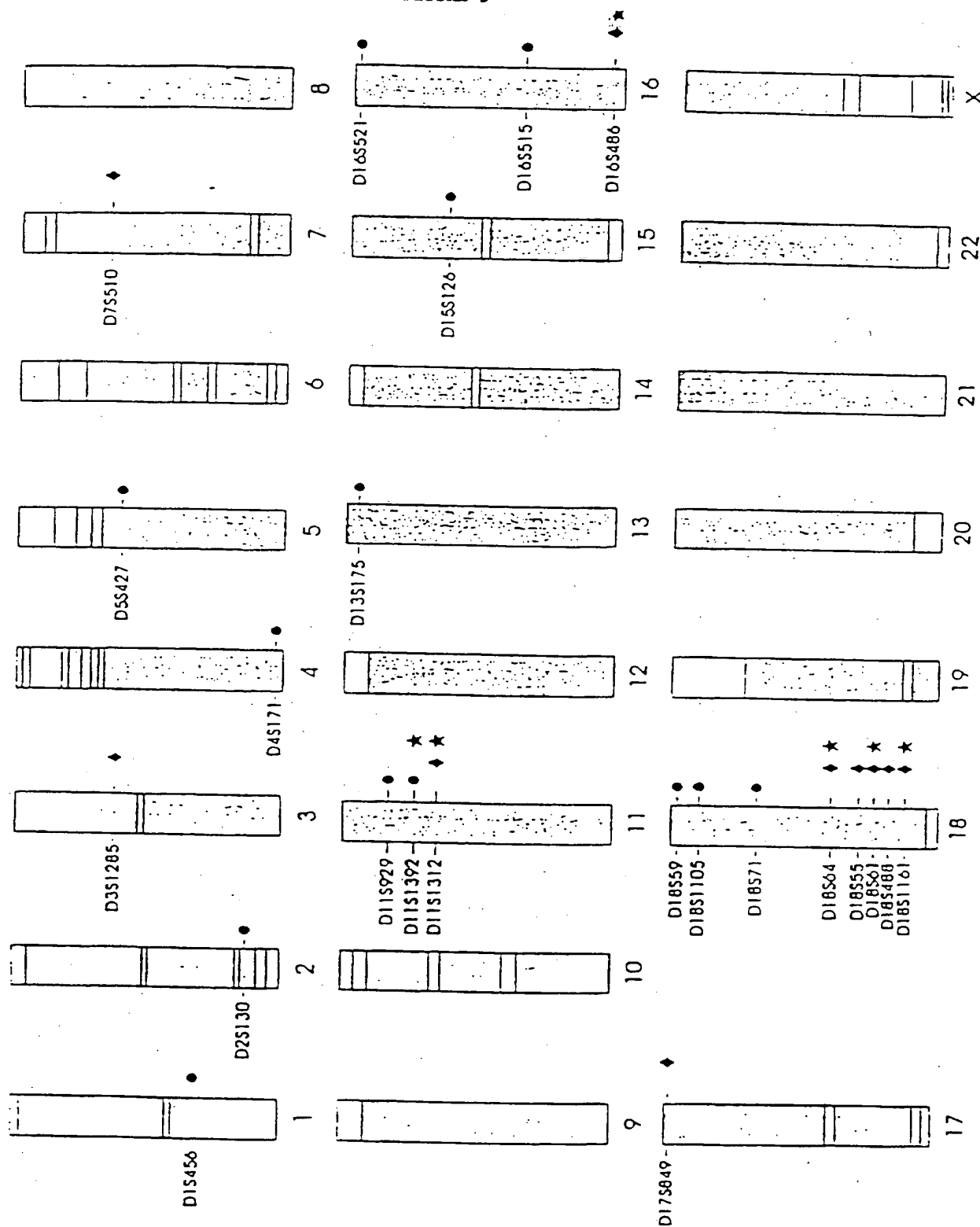
Figure 2

Table 1. Lod scores for markers exceeding the arbitrary coverage thresholds.

Marker Name	distance from pter	Family CR001		Family CR004		Combined	
		$Z_{\text{max}}$ $\geq 0.8$	Theta	$Z_{\text{max}}$ $\geq 1.2$	Theta	$Z_{\text{max}}$ $\geq 1.6$	Theta
D1S456	224.6	1.32	0.0	0.0	0.50	0.0	0.50
D2S130	230.1	0.89	0.0	0.12	0.35	0.36	0.26
D3S1285	91.0	0.00	0.50	2.59	0.00	1.15	0.16
D4S171	207.9	1.07	0.07	0.01	0.05	0.22	0.29
D5S427	69.6	1.39	0.0	0.0	0.50	0.7	0.18
D7S510	60.5	0.04	0.40	2.04	0.0	0.82	0.17
D11S929	36.3	0.80	0.11	0.03	0.42	0.43	0.24
D11S1392	38.6	0.86	0.07	0.90	0.23	1.58	0.19
D11S1312	42.0	0.47	0.13	1.77	0.0	1.95	0.05
D13S175	7.4	0.83	0.0	0.0	0.50	0.24	0.15
D13S126	45.5	1.09	0.0	0.0	0.48	0.06	0.40
D16S521	4.6	1.46	0.0	0.41	0.26	1.18	0.17
D16S515	94.8	0.93	0.09	0.01	0.46	0.39	0.23
D16S486	133.6	0.27	0.19	1.29	0.20	1.60	0.20
D17S849	0.60	0.0	0.50	1.22	0.07	0.32	0.14
D18S59	1.1	1.43	0.0	0.0	0.50	0.02	0.46
D18S1105	2.8	0.97	0.0	0.01	0.47	0.01	0.46
D18S71	43.8	0.96	0.0	0.0	0.50	0.0	0.50
D18S64	84.0	0.33	0.11	1.34	0.15	1.67	0.13
D18S55	95.5	0.0	0.50	2.09	0.13	1.51	0.18
D18S61	103.8	0.0	0.50	2.26	0.12	1.94	0.16
D18S488	105.6	0.0	0.50	1.26	0.14	1.02	0.19
D18S1161	113.0	0.0	0.50	1.79	0.16	1.76	0.17

Markers for which lod scores exceeded the arbitrary thresholds used for genome coverage calculations (in bold).  $Z_{\text{max}}$  is the maximum likelihood estimate of the lod score at the corresponding value of the recombination fraction (theta).

FIGURE 3



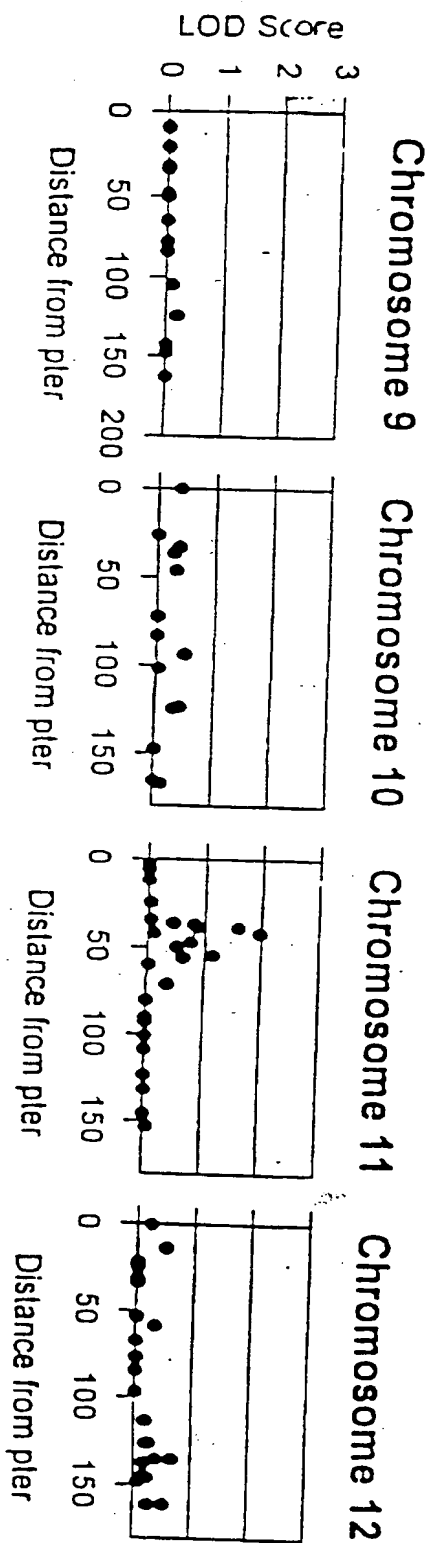
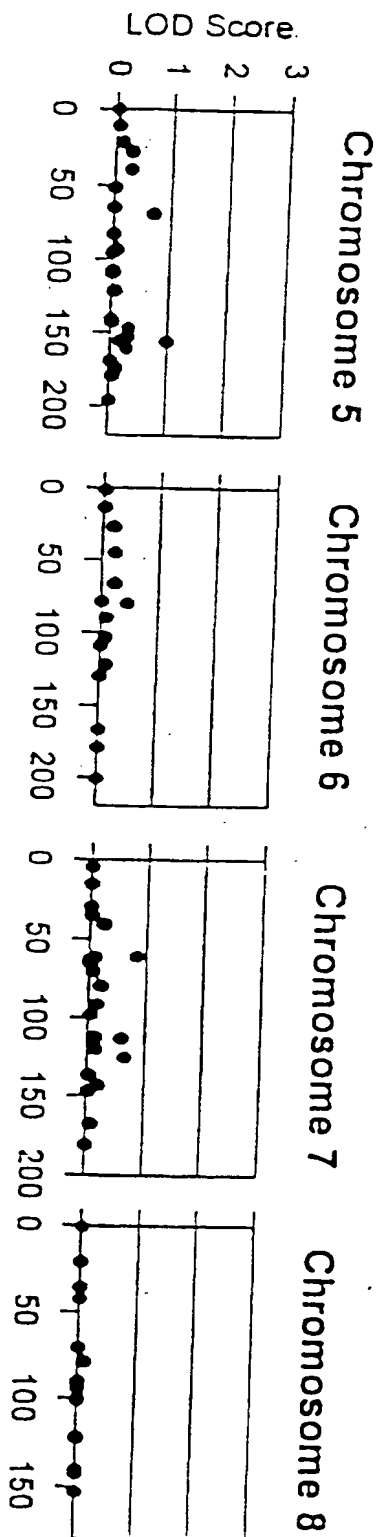
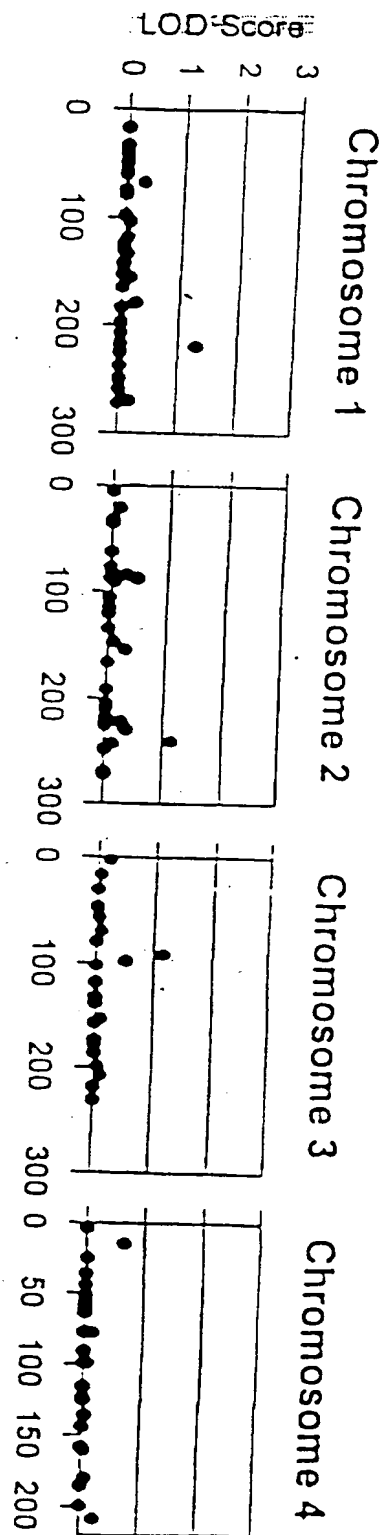


FIGURE 4b

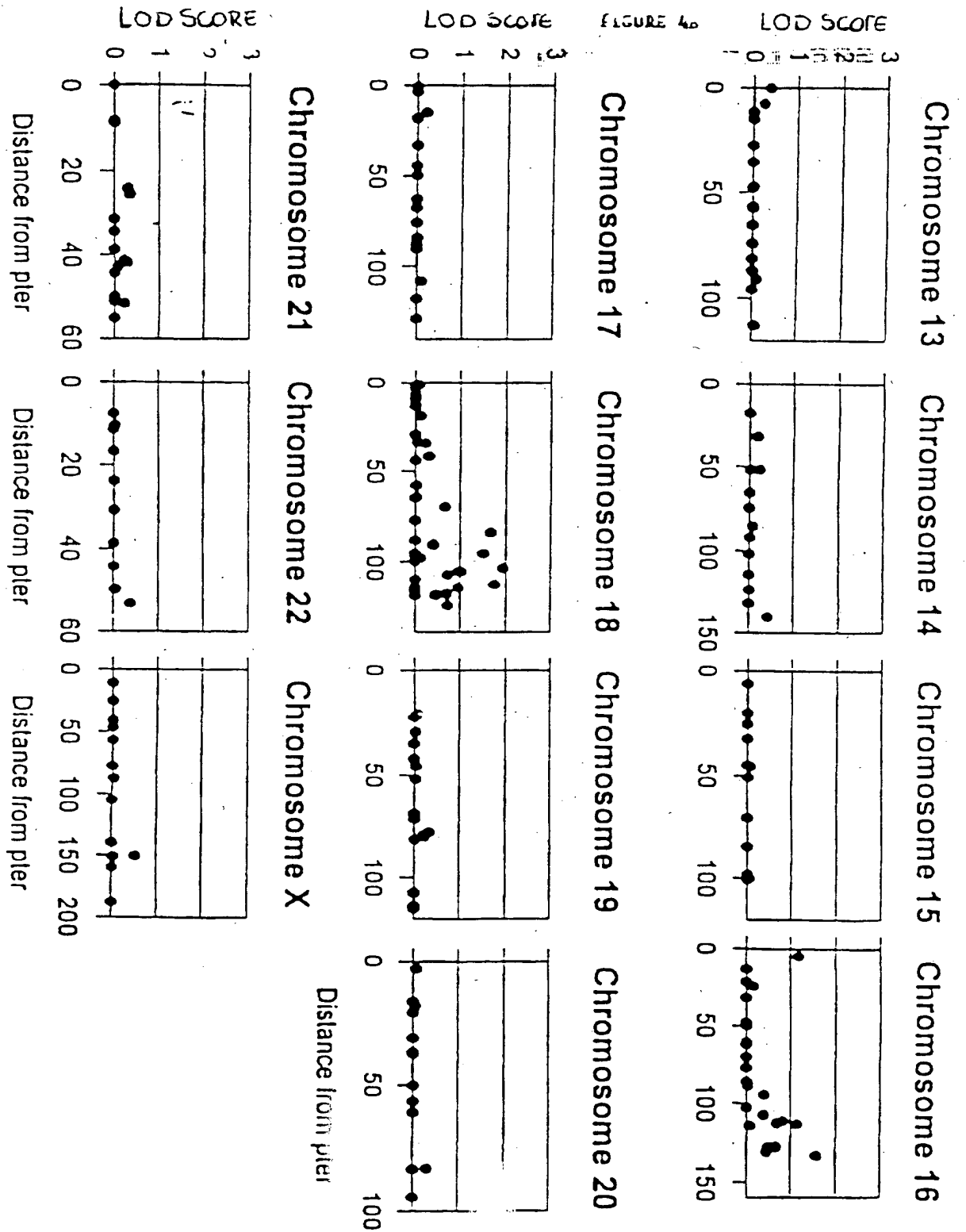
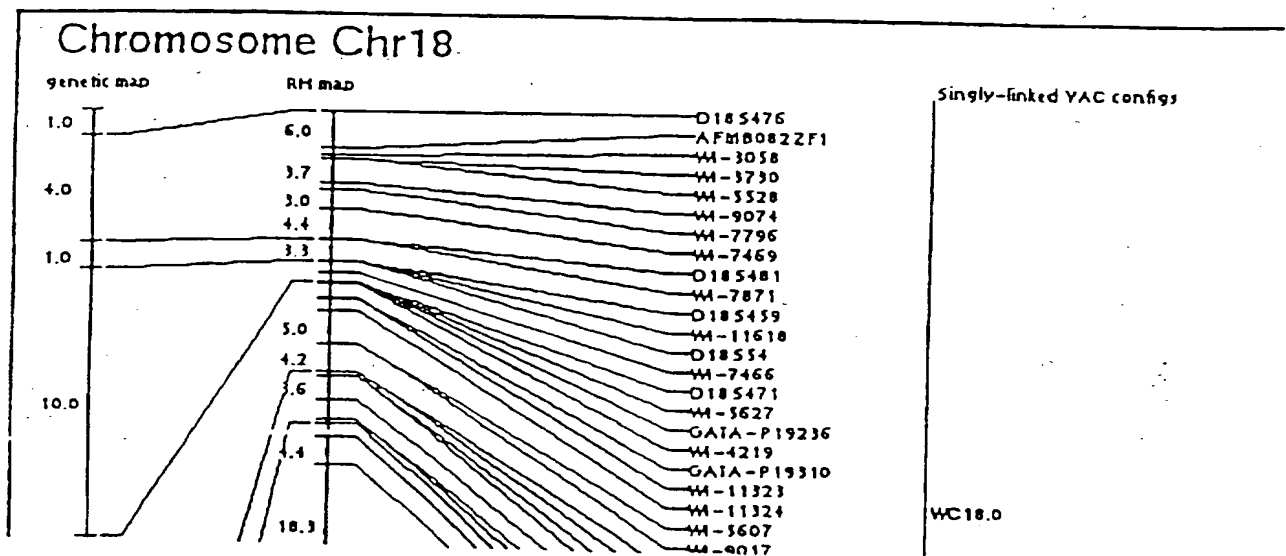


FIGURE 5

## Chr18: Contigs Anchored on Integrated Map

Be patient... This is a large image!



You can click on the name of an STS or a contig in order to retrieve information about it. [Download this map](#) as a PICT file (Macintosh) or a GIF (everybody else)

### NOTES

1. This is a composite map in which the genetic linkage map from Génethon, and the radiation hybrid map from the Whitehead Institute/MIT Center are used to anchor YAC/STS contigs. We only show the subset of genetic- and radiation-hybrid mapped STSs for which positive YACs are present. For the genetic map, please refer to the linkage maps published in *Nature Genetics* 7(2):246-339 (1994) for the complete genetic maps.
2. The apparent size of a contig on this map does not always correlate with the number of its members. Some apparent "large" contigs are artificially expanded because of contradictions between the radiation hybrid map position of one or more markers on the genetic map, and adjacencies computed from YAC linkage. Contigs that appear to overlap may represent places where missing YAC data prevents the contigs from merging, or, in some cases, contradictions between the order derived from the radiation hybrid map and the order derived from the STS content map.
3. The large central gap that appears on many of the radiation hybrid maps corresponds to the centromere.
4. Markers derived from expressed sequence tags (ESTs) or other expressed sequences are colored red.

FIGURE 6A

This STS is part of singly-linked contig WC18.0:

	STS	Map Position		Contig	
		Chrom	Genetic	RH	Single Double
1	<u>WI-9527</u>	<u>Chr18</u>	-	-	<u>WC18.0</u> <u>WC-1465</u>
2	<u>CHLC.GGAT2G04</u>	<u>Chr18</u>	-	-	<u>WC18.0</u> <u>WC-1465</u>
3	<u>CHLC.GGAT2G04.1217</u>	<u>Chr18</u>	-	-	<u>WC18.0</u> <u>WC-1465</u>
4	<u>D18S59</u>	<u>Chr18</u>	<u>0 cM</u>	-	<u>WC18.0</u> <u>WC-1465</u>
5	<u>D18S1140</u>	<u>Chr18</u>	<u>0 cM</u>	-	<u>WC18.0</u> <u>WC-1465</u>
6	<u>WI-7796</u>	<u>Chr18</u>	-	<u>15 cR</u>	<u>WC18.0</u> -
7	<u>WI-9074</u>	<u>Chr18</u>	-	<u>12 cR</u>	<u>WC18.0</u> <u>WC-1465</u>
8	<u>WI-5528</u>	<u>Chr18</u>	-	<u>7 cR</u>	<u>WC18.0</u> -
9	<u>D18S476</u>	<u>Chr18</u>	<u>1 cM</u>	<u>0 cR</u>	<u>WC18.0</u> -
10	<u>WI-7226</u>	<u>Chr18</u>	-	-	<u>WC18.0</u> <u>WC-909</u>
11	<u>AFMB324ZE5</u>	<u>Chr18</u>	-	-	<u>WC18.0</u> <u>WC-909</u>
12	<u>AFMB082ZF1</u>	<u>Chr18</u>	-	<u>5 cR</u>	<u>WC18.0</u> <u>WC-909</u>
13	<u>D18S1146</u>	<u>Chr18</u>	<u>1 cM</u>	-	<u>WC18.0</u> <u>WC-909</u>
14	<u>WI-3058</u>	<u>Chr18</u>	-	<u>5 cR</u>	<u>WC18.0</u> <u>WC-909</u>
15	<u>D18S1105</u>	<u>Chr18</u>	<u>1 cM</u>	-	<u>WC18.0</u> <u>WC-909</u>
16	<u>WI-3730</u>	<u>Chr18</u>	-	<u>5 cR</u>	<u>WC18.0</u> <u>WC-1576</u>
17	<u>AFM077YD11</u>	<u>Chr18</u>	-	-	<u>WC18.0</u> <u>WC-1576</u>
18	<u>D18S1098</u>	<u>Chr18</u>	<u>4 cM</u>	-	<u>WC18.0</u> <u>WC-1576</u>
19	<u>WI-7469</u>	<u>Chr18</u>	-	<u>16 cR</u>	<u>WC18.0</u> <u>WC-1576</u>
20	<u>WI-7871</u>	<u>Chr18</u>	-	<u>22 cR</u>	<u>WC18.0</u> <u>WC-1576</u>
21	<u>D18S481</u>	<u>Chr18</u>	<u>5 cM</u>	<u>21 cR</u>	<u>WC18.0</u> <u>WC-1576</u>
22	<u>WI-4747</u>	<u>Chr18</u>	-	-	<u>WC18.0</u> <u>WC-1576</u>
23	<u>D18S1154</u>	<u>Chr18</u>	<u>6 cM</u>	-	<u>WC18.0</u> <u>WC-1576</u>
24	<u>CHLC.ATA14B09</u>	<u>Chr18</u>	-	-	<u>WC18.0</u> <u>WC-1576</u>
25	<u>WI-7466</u>	<u>Chr18</u>	-	<u>27 cR</u>	<u>WC18.0</u> <u>WC-1576</u>
26	<u>D18S54</u>	<u>Chr18</u>	<u>6 cM</u>	<u>25 cR</u>	<u>WC18.0</u> <u>WC-1576</u>
27	<u>D18S63</u>	<u>Chr18</u>	<u>6 cM</u>	-	<u>WC18.0</u> <u>WC-1576</u>
28	<u>D18S459</u>	<u>Chr18</u>	<u>6 cM</u>	<u>25 cR</u>	<u>WC18.0</u> <u>WC-1576</u>
29	<u>WI-6014</u>	<u>Chr18</u>	-	-	<u>WC18.0</u> <u>WC-1576</u>
30	<u>WI-4219</u>	<u>Chr18</u>	-	<u>28 cR</u>	<u>WC18.0</u> <u>WC-143</u>
31	<u>AFM238YG3</u>	<u>Chr18</u>	-	-	<u>WC18.0</u> <u>WC-143</u>
32	<u>D18S471</u>	<u>Chr18</u>	<u>17 cM</u>	<u>28 cR</u>	<u>WC18.0</u> <u>WC-143</u>
33	<u>D18S458</u>	<u>Chr18</u>	<u>17 cM</u>	-	<u>WC18.0</u> <u>WC-143</u>

FIGURE 6B

34	<u>D18S452</u>	<u>Chr18</u>	<u>17 cM</u>	-	<u>WC18.0</u>	<u>WC-143</u>
35	<u>D18S62</u>	<u>Chr18</u>	<u>17 cM</u>	-	<u>WC18.0</u>	<u>WC-143</u>
36	<u>WI-5627</u>	<u>Chr18</u>	-	<u>28 cR</u>	<u>WC18.0</u>	<u>WC-143</u>
37	<u>CHLC.GATA82D03</u>	<u>Chr18</u>	-	<u>28 cR</u>	<u>WC18.0</u>	<u>WC-143</u>
38	<u>FB25F12</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-143</u>
39	<u>CHLC.GATA51H07</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-143</u>
40	<u>CHLC.GATA88A12</u>	<u>Chr18</u>	-	<u>30 cR</u>	<u>WC18.0</u>	<u>WC-143</u>
41	<u>WI-9619</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-143</u>
42	<u>AFMB346YA9</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-143</u>
43	<u>AFM323TC9</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-862</u>
44	<u>WI-5607</u>	<u>Chr18</u>	-	<u>36 cR</u>	<u>WC18.0</u>	<u>WC-862</u>
45	<u>WI-9017</u>	<u>Chr18</u>	-	<u>36 cR</u>	<u>WC18.0</u>	<u>WC-862</u>
46	<u>AFM077YF7</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-934</u>
47	<u>WI-8546</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-934</u>
48	<u>CHLC.GGAA16G02</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-934</u>
49	<u>D18S464</u>	<u>Chr18</u>	<u>32 cM</u>	<u>46 cR</u>	<u>WC18.0</u>	<u>WC-934</u>
50	<u>NIB1802</u>	<u>Chr18</u>	-	<u>56 cR</u>	<u>WC18.0</u>	<u>WC-934</u>
51	<u>D18S1153</u>	<u>Chr18</u>	<u>34 cM</u>	-	<u>WC18.0</u>	<u>WC-934</u>
52	<u>D18S1150</u>	<u>Chr18</u>	<u>36 cM</u>	-	<u>WC18.0</u>	<u>WC-934</u>
53	<u>WI-4589</u>	<u>Chr18</u>	-	<u>58 cR</u>	<u>WC18.0</u>	<u>WC-934</u>
54	<u>WI-4319</u>	<u>Chr18</u>	-	<u>62 cR</u>	<u>WC18.0</u>	<u>WC-934</u>
55	<u>D18S1158</u>	<u>Chr18</u>	<u>38 cM</u>	-	<u>WC18.0</u>	<u>WC-934</u>
56	<u>D18S1116</u>	<u>Chr18</u>	<u>40 cM</u>	-	<u>WC18.0</u>	<u>WC-377</u>
57	<u>CHLC.GATA11A06.668</u>	<u>Chr18</u>	-	<u>48 cR</u>	<u>WC18.0</u>	<u>WC-377</u>
58	<u>CHLC.GATA11A06</u>	<u>Chr18</u>	-	<u>54 cR</u>	<u>WC18.0</u>	<u>WC-377</u>
59	<u>D18S53</u>	<u>Chr18</u>	<u>41 cM</u>	-	<u>WC18.0</u>	<u>WC-377</u>
60	<u>WI-9134</u>	<u>Chr18</u>	-	<u>52 cR</u>	<u>WC18.0</u>	<u>WC-377</u>
61	<u>IB1114</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-377</u>
62	<u>D18S482</u>	<u>Chr18</u>	<u>41 cM</u>	<u>56 cR</u>	<u>WC18.0</u>	<u>WC-377</u>
63	<u>WI-2382</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-377</u>
64	<u>WI-6819</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-377</u>
65	<u>D18S71</u>	<u>Chr18</u>	<u>43 cM</u>	<u>84 cR</u>	<u>WC18.0</u>	<u>WC-377</u>
66	<u>AFMA058YG5</u>	<u>Chr18</u>	-	<u>80 cR</u>	<u>WC18.0</u>	<u>WC-377</u>
67	<u>WI-5506</u>	<u>Chr18</u>	-	<u>90 cR</u>	<u>WC18.0</u>	<u>WC-377</u>
68	<u>D18S453</u>	<u>Chr18</u>	<u>43 cM</u>	<u>93 cR</u>	<u>WC18.0</u>	<u>WC-738</u>
69	<u>D18S73</u>	<u>Chr18</u>	<u>43 cM</u>	-	<u>WC18.0</u>	<u>WC-377</u>
70	<u>STSG-10174</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-377</u>

FIGURE 6C

71	<u>CHLC.GCT5D07</u>	<u>Chr18</u>	-	<u>101 cR</u>	<u>WC18.0</u>	<u>WC-377</u>
72	<u>WI-10768</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1182</u>
73	<u>D18S1149</u>	<u>Chr18</u>	<u>49 cM</u>	-	<u>WC18.0</u>	<u>WC-1182</u>
74	<u>WI-1869</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1182</u>
75	<u>D18S1104</u>	<u>Chr18</u>	<u>49 cM</u>	-	<u>WC18.0</u>	<u>WC-1182</u>
76	<u>AFMA205YH5</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1182</u>
77	<u>AFMB340VE5</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1182</u>
78	<u>CHLC.GATA41G05</u>	<u>Chr18</u>	-	<u>185 cR</u>	<u>WC18.0</u>	<u>WC-1182</u>
79	<u>AFMB319WF9</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1182</u>
80	<u>D18S44</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1182</u>

Details on contig assembly.

---

18p allele frequencies

Fig. 7

MARKERNAME		aff 105	ntrans	control	
D18SAVA5	225	0.04	0.02		
	227	0.29	0.24		
	229	0.22	0.15		
	231	0.04	0.08		
	233	0.14	0.23		
	235	0.25	0.22		
	237	0.02	0.03		
	239	0.00	0.00		
D18SCA211	183	0.02	0.04	0.01	
	189	0.00	0.01	0.01	
	191	0.01	0.00	0.03	
	193	0.24	0.17	0.33	
	195	0.21	0.19	0.18	
	197	0.06	0.11	0.03	
	199	0.06	0.04	0.01	
	201	0.10	0.14	0.10	
	203	0.02	0.04	0.06	
	205	0.16	0.18	0.14	
	207	0.09	0.04	0.06	
	209	0.02	0.02	0.02	
D18SCA212	211	0.01	0.00	0.00	
	215	0.00	0.00	0.00	
	217	0.00	0.00	0.01	
	200	0.40	0.40	0.39	
	202	0.31	0.32	0.29	
	204	0.05	0.05	0.03	
	206	0.04	0.06	0.10	
	214	0.01	0.00	0.00	
	216	0.14	0.12	0.15	

18p allele frequencies

MARKERNAME		aff 105	ntrans	control		
	218	0.04	0.00	0.04		
D18S1140	256	0.06	0.07	0.06		
	268	0.77	0.72	0.73		
	270	0.02	0.00	0.06		
	272	0.03	0.03	0.03		
	274	0.00	0.00	0.00		
	276	0.03	0.06	0.02		
	278	0.02	0.06	0.05		
	280	0.04	0.06	0.02		
	282	0.01	0.00	0.02		
MARKERNAME		aff 105	ntrans	control		
D18S59	148	0.16	0.26	0.21		
	150	0.07	0.09	0.14		
	152	0.02	0.06	0.01		
	154	0.36	0.19	0.28	0.17	0.08
	156	0.04	0.04	0.08		
	158	0.22	0.21	0.13		
	160	0.04	0.08	0.05		
	162	0.05	0.06	0.05		
	164	0.02	0.01	0.02		
	168	0.00	0.00	0.01		
D18STA201	214	0.02	0.00	0.00		
	220	0.09	0.09	0.04		
	222	0.01	0.00	0.01		
	228	0.01	0.01	0.00		
	230	0.25	0.22	0.16	0.03	0.09

MARKERNAME	aff 105	ntrans	control
232	0.07	0.04	0.07
234	0.02	0.00	0.00
236	0.01	0.00	0.00
238	0.01	0.00	0.00
242	0.09	0.09	0.04
244	0.13	0.13	0.19
246	0.09	0.09	0.11
248	0.06	0.11	0.10
250	0.07	0.07	0.06
252	0.07	0.10	0.12
254	0.02	0.03	0.03
256	0.01	0.01	0.03
258	0.01	0.01	0.01
260	0.01	0.09	0.02
262	0.01	0.00	0.00
D18CA231			
182	0.00	0.00	0
184	0.20	0.23	0.26
186	0.70	0.66	0.68
188	0.00	0.01	0.01
190	0.02	0.00	0.02
192	0.00	0.00	0.01
194	0.02	0.02	0
196	0.00	0.00	0
198	0.02	0.01	0
200	0.01	0.01	0.01
202	0.02	0.03	0.01
MARKERNAME			
	aff 105	ntrans	control

18p allele frequencies

MARKERNAME		af 105	ntrans	control		
D18SAT201	170	0.53	0.55	0.52		
	174	0.00	0.01	0.01		
	178	0.37	0.36	0.36		
	182	0.01	0.00	0.00		
	186	0.07	0.06	0.07		
	190	0.01	0.00	0.00		
	194	0.01	0.01	0.03		
D18SCA225	160	0.16	0.20	0.21		
	168	0.02	0.04	0.00		
	170	0.00	0.00	0.01		
	172	0.47	0.38	0.42	0.09	0.04
	174	0.22	0.24	0.26		
	176	0.04	0.04	0.05		
	178	0.04	0.04	0.02		
	180	0.02	0.01	0.01		
	184	0.03	0.00	0.02		
D18SW3442	11	0.42	0.28	0.36	0.14	0.06
	12	0.01	0.01	0.01		
	14	0.07	0.11	0.11		
	16	0.12	0.17	0.12		
	18	0.18	0.15	0.14		
	20	0.05	0.09	0.09		
	22	0.08	0.10	0.11		
	24	0.05	0.08	0.03		
	26	0.00	0.00	0.02		
	38	0.00	0.00	0.00		
D18SCA213	112	0.12	0.17	0.07		
	120	0.00	0.05	0.01		
	122	0.03	0.03	0.04		
	124	0.44	0.37	0.46		

18p allele frequencies

MARKERNAME		aff 105	ntrans	control		
	126	0.30	0.24	0.35		
	128	0.08	0.11	0.06		
	130	0.00	0.00	0.00		
	132	0.03	0.02	0.01		
D18SGAT201	142	0.04	0.06	0.02		
	146	0.08	0.08	0.06		
	150	0.61	0.62	0.69		
	154	0.15	0.15	0.12		
	158	0.11	0.07	0.10		
	162	0.02	0.02	0.00		
D18SGAT203						
	188	0.42	0.37	0.38		
	192	0.12	0.14	0.17		
	196	0.01	0.04	0.01		
	200	0.02	0.04	0.01		
	204	0.06	0.02	0.04		
	208	0.19	0.21	0.20		
	212	0.11	0.11	0.11		
	216	0.09	0.07	0.08		
D18SCA219	221	0.00		0.01		
	223	0.00		0.00		
	225	0.00		0.00		
	233	0.00		0.00		
	235	0.22		0.21		
	239	0.02		0.01		
	241	0.54		0.63		
	243	0.07		0.07		
	245	0.13		0.06		
MARKERNAME		aff 105	ntrans	control		

## 18p allele frequencies

MARKERNAME		aff 105	ntrans	control	
D18S1105	101	0.16	0.11		
	103	0.12	0.08		
	105	0.03	0.02		
	81	0.02	0.01		
	83	0.01	0.02		
	85	0.51	0.54		
	87	0.01	0.06		
	91	0.00	0.00		
	95	0.01	0.04		
D18SCA209	97	0.04	0.04		
	99	0.06	0.06		
	173	0.57	0.53	0.69	
	175	0.02	0.03	0.04	
	177	0.20	0.18	0.09	
	179	0.01	0.03	0.00	
	181	0.19	0.24	0.18	
	187	0.00	0.00	0.00	
D18SCA202	182	0.16	0.14		
	184	0.02	0.00		
	186	0.01	0.01		
	190	0.09	0.02		
	192	0.10	0.16		
	194	0.10	0.09		
	196	0.37	0.35		
	198	0.09	0.10		
	200	0.05	0.08		
D18S1146	202	0.00	0.03		
	208	0.00	0.00		
	270	0.32	0.35		
	272	0.07	0.10		
	274	0.60	0.51		

18p allele frequencies

MARKERNAME	276	aff 105	ntrans	control		
		0.02	0.04			
D18S166D05	300	0.17	0.21	0.19		
	304	0.16	0.12	0.14		
	308	0.18	0.18	0.13		
	312	0.35	0.26	0.36	..	
	316	0.08	0.18	0.11		
	320	0.04	0.04	0.03		
	324	0.01	0.01	0.02		
D18S476	261	0.00	0.01	0.01		
	263	0.01	0.04	0.04		
	265	0.05	0.12	0.04		
	267	0.20	0.26	0.23		
	269	0.08	0.09	0.04		
	271	0.56	0.38	0.54	...	
	273	0.04	0.08	0.07		
	275	0.04	0.03	0.03		

Fig 8

Affected haplotypes

18p	PAN	MAN	ca212	1140	59	ca231	ta201	al201	PDca225	w3442	ca213	ga201	iga203
200	279	280	218	1268	1158	1186	1242	1178	1160	114	1112	1150	1188
200			200	1268	1158	1186	1248	1170	1160	114	1124	1150	1208
204	309	349	200	1282	1150	1202	1220	1170	1160	118	1124	1154	1208
204			206	1268	1158	1184	1250	1170	1184	122	1112	1158	1212
206	1	2	218	0276	0156	0186	0252	0186	0174	018	0124	1150	0212
206			200	0268	0148	0184	0248	0170	0160	014	0124	1146	0188
207	277	278	200	1268	1154	1194	1220	1170	1178	118	1128	1146	1192
207			204	1268	1158	1184	1230	1178	1176	122	1112	1154	1216
209	0	0	200	1268	0154	1186	0242	1170	1172	110	1126	0146	1188
209			200	1256	0150	1184	0254	1186	1172	116	1124	0158	1188
213	0	0	216	0272	1150	1186	0250	1170	1160	110	1124	1150	1212
213			200	0282	1150	1184	0238	1170	1180	114	1124	1150	1196
214	460	459	202	1268	1158	1200	1220	1178	1176	110	1126	1150	1188
214			216	1276	1154	1186	1242	1170	1176	118	1124	1150	1188
215	1	270	218	1276	1160	0186	1242	1178	1174	114	1124	1154	0192
215			200	1268	1154	0186	1230	1170	1160	114	1124	1150	0188
216	1	259	204	1278	0156	1186	0230	1178	1170	116	1130	1154	0216
216			200	1268	0162	1184	0252	1170	1160	114	1128	1150	0192
218	273	272	200	1268	1162	1186	1220	1186	1172	010	1112	1150	1212
218			200	1268	1158	1186	1246	1170	1174	020	1124	1158	1188
219	0	0	202	1256	1154	1186	1230	1178	1172	110	1124	1154	1188
219			200	1268	1168	1184	1250	1170	1174	116	1126	1146	1188
220	267	2	216	0268	1152	1186	1230	1178	1176	110	1126	1154	1208
220			200	0268	1154	1186	1232	1178	1172	110	1126	1142	1212
221	0	0	202	1268	1160	1184	0250	1178	0174	018	0124	1154	1216
221			202	1268	1154	1186	0250	1170	0172	010	0126	1158	1188
223	0	0	202	1280	0	148	0256	0186	0174	018	1124	1158	0212
223			202	1268	0	154	0252	0178	0172	018	1124	1146	0208
225	264	2	200	1268	1164	1186	1230	1178	0172	026	0124	1158	1216
225			200	1268	1158	1186	1246	1170	0172	010	0124	1158	1188
226	1	2	202	1268	0154	0186	0	0178	0	010	0124	1150	0188
226			202	1256	0148	0184	0	0170	0	172	0124	1142	0188
228	1	260	200	1268	1150	1202	1220	1170	1174	018	1128	1150	1192
228			200	1268	1158	1186	1242	1178	1172	018	1124	1158	1208

## Affected haplotypes

18p	PAN	MAN	ca212	1140	59	ca231	la201	al201	PDca225	w3442	ca213	ga201	ga203
229	257	2	200	1 268	0 154	1 186	1 244	1 170	1 174	1 10	1 126	1 150	1 192
229			216	1 256	0 158	1 186	1 244	1 186	1 174	1 24	1 124	1 146	1 216
230	0	0	202	1 268	1 160	1 186	1 230	1 170	1 172	0 18	1 122	1 150	1 208
230			202	1 268	1 158	1 186	1 248	1 170	1 160	0 12	1 124	1 150	1 216
231	299	298	216	1 268	1 158	1 186	1 220	1 170	1 172	1 20	1 124	1 150	1 204
231			218	1 268	1 158	1 186	1 244	1 170	1 174	1 22	1 126	1 150	1 204
232	1	310	206	1 268	1 150	1 186	1 222	1 170	1 172	1 20	1 124	1 154	0 188
232			200	1 268	1 158	1 186	1 230	1 170	1 178	1 10	1 126	1 150	0 188
233			200	1 268	1 148	1 184	1 252	1 170	1 174	1 10	1 126	1 162	1 208
234	1	261	200	1 268	1 158	1 186	1 262	1 170	1 174	1 24	1 126	1 150	1 192
234			200	1 268	0 150	1 186	0 248	1 170	1 172	1 10	1 112	1 154	0 192
235	0	0	200	1 276	0 150	1 186	0 248	1 170	1 174	1 22	1 124	1 150	0 192
235			202	1 268	0 156	1 184	0 214	1 170	1 174	1 16	1 126	0 150	1 208
237	0	0	200	1 268	1 158	1 186	1 214	1 178	1 172	1 16	1 124	0 154	1 208
237			200	1 268	1 154	1 186	1 230	1 186	1 172	1 16	1 128	1 150	1 208
238	456	457	202	1 268	1 154	1 186	1 230	1 178	1 172	1 10	1 112	1 150	1 208
238			200	1 268	1 158	1 186	1 230	1 170	1 178	1 14	1 112	1 150	1 188
239	312	2	218	1 268	1 160	1 186	0 248	1 170	1 172	1 16	1 124	1 154	0 208
239			200	1 268	1 158	1 184	0 242	1 178	1 172	1 18	1 124	1 150	0 188
240	1	2	200	1 268	1 158	0 186	1 242	0 178	1 172	1 18	1 128	0 154	0 188
240			200	1 268	1 148	0 186	1 230	0 178	1 172	1 18	1 124	0 146	0 188
241	1	342	216	1 268	1 158	1	0 246	1 170	1 172	1 20	1 126	0 150	1 188
241			200	1 268	1 158	1	0 250	1 170	1 172	1 10	1 124	0 142	1 188
242	0	0	216	1 268	1 156	0 186	1	1 186	1 174	1 14	0 126	1 150	1 192
242			200	1 268	1 154	0 186	1	1 170	1 160	1 10	0 126	1 150	1 188
243	347	274	200	1 268	1 154	1 186	1 230	1 178	0 172	0 10	1 124	1 150	1
243			218	1 268	1 150	1 186	1 252	1 170	0 160	0 38	1 124	1 146	1
245	0	0	200	1 268	1 154	1 186	1 232	1 178	0 172	1 10	1 126	1 154	1 216
245			202	1 268	1 150	1 186	1 242	1 170	0 172	1 16	1 124	1 150	1 192
246	1	262	204	0 270	1 158	1 186	1 246	1 178	0 172	1 16	1 126	1 150	1 188
246			202	0 268	1 154	1 186	1 242	1 170	0 172	1 22	1 122	1 150	1 216
247	303	302	202	1 268	1 154	1 186	1 230	1 178	1 174	1 10	1 124	1 158	1 188
247			200	1 268	1 154	1 186	1 242	1 170	1 176	1 10	1 126	1 150	1 216
248	334	333	200	1 268	1 154	1 184	1 232	1 170	1 160	1 20	1 112	1 150	1
248			202	1 268	1 154	1 186	1 244	1 170	1 174	1 16	1 112	1 146	1

## Affected haplotypes

18p	PAN	MAN	ca212	1140	59	ca231	ta201	ai201	PDca225	w3442	ca213	ga201	ga203
249	1	2	200	0 268	0 154	0 186	1	230 0 194	0	172 0	10 0 124	1 150	1 188
249			216	0 256	0 148	0 186	1	246 0 178	0	174 0	16 0 124	1 150	1 188
251	301	300	216	1 272	1 150	1 184	1 250	1 170	1 160	1 110	1 124	1 150	1 212
251			216	1 268	1 158	1 186	1 244	1 186	1 174	1 120	1 124	1 150	1 188
252	1	285	200	0 268	1 154	1 186	1 230	1 178	1 172	1 110	1 124	0 150	1 188
252			204	0 268	1 158	1 186	1 246	1 170	1 160	1 118	1 126	0 150	1 216
253	1	258	216	0 268	1 160	1 186	1 228	1 170	1 160	1 116	1 124	1 150	1 188
253			200	0 268	1 154	1 186	1 230	1 178	1 160	1 116	1 126	1 150	1 216
254	467	2	202	1 268	1 160	1 186	1 230	1 178	1 170	1 118	1 122	1 150	1 208
254			200	1 268	1 154	1 186	1 230	1 178	1 178	1 110	1 124	1 142	1 188
265	1	266	216	1 272	1 150	1 184	1 250	1 170	1 160	1 110	1 126	0 150	1 212
265			202	1 268	1 154	1 186	1 230	1 178	1 172	1 110	1 124	0 150	1 188
311	1	485	216	1 268	1 154	1 186	1 244	1 170	1 160	1 110	1 126	1 150	1 188
311			200	1 268	1 162	1 186	1 242	1 186	1 174	1 110	1 124	1 158	1 208
314	348	313	200	1 268	1 148	1 184	1 248	1 170	1 168	1 118	0 124	1 150	1 208
314			216	1 268	1 162	1 184	1 250	1 170	1 172	1 110	0 126	1 150	1 188
316	1	317	214	1 268	1 154	1 186	1 230	1 178	1 172	1 110	1 124	1 150	1 208
316			200	1 268	1 154	1 186	1 242	1 170	1 172	1 110	1 126	1 150	1 188
319	318	2	202	0 272	0 158	0 184	0 244	1 178	0 184	0 10	1 126	0 150	1 188
319			200	0 256	0 154	0 186	0 244	1 170	0 174	0 10	1 112	0 150	1 188
321	1	320	202	0 268	1 158	0 186	0	0 178	1 178	0	1 128	0	0
321			200	0 268	1 154	0	0	0 170	1 172	0	1 124	0	0
324	0	0	202	1 268	1 158	1 186	1 232	1 178	1 172	0 24	1 112	1 150	1 212
324			216	1 268	1 150	1 196	1 220	1 170	1 160	0 18	1 128	1 154	1 208
326	325	336	206	1 280	1 152	1 198	1 232	1 170	1 172	1 16	1 124	1 150	1 188
326			202	1 268	1 154	1 186	1 232	1 178	1 172	1 16	1 132	1 150	1 192
329	1	330	200	1 268	1 154	0 186	1 248	1 170	1 160	1 14	1 128	1 150	1 188
329			206	1 268	1 148	0 186	1 234	1 170	1 172	1 22	1 124	1 150	1 208
211	1	2	200	0 268	1 154	0	186 0	230 0 178	1 172	1	10 0 126	0	150
211			204	0 268	1 148	0	198 0	252 0 178	1 172	1	18 0 112	0	154
353	1	352	218	1 280	0 148	1 186	1 246	1 170	1 160	1 18	1 132	1 154	1 192
353			200	1 268	0 148	1 186	1 246	1 170	1 172	1 18	1 112	1 146	1 192
356	362	2	216	1 268	1 154	1	186 0 248	1 178	0 172	1	10 0 124	1 150	1 208
356			204	1 268	1 164	1	190 0 232	1 170	0 172	1	18 0 126	1 150	1 216

## Affected haplotypes

18p	PAN	MAN	ca212	1140	59	ca231	la201	at201	PDca225	w3442	ca213	ga201	ga203
357	1	358	202	0 268	1	154	0 186	1 232	1 178	1 160	1 128	0 150	1 196
357			214	0 278	1	158	0 186	1 248	1 178	1 184	1 124	0 150	1 208
359	378	365	202	1 268	1 154	1 186	1 186	1 230	1 178	1 172	1 126	1 154	1 188
359			202	1 272	1 158	1 184	1 184	1 244	1 178	1 184	1 112	1 150	1 188
367	1	366	202	1 268	1 154	1 186	1 186	1 232	1 178	1 172	1 126	1 158	0 208
367			202	1 268	1 154	1 186	1 186	1 242	1 178	1 172	1 112	1 142	0 208
372	1	370	200	1 268	1 154	1 186	0	0	0 172	1 110	1	124	0 150
372			216	1 268	1 148	1 184	0	0	0 174	1 110	1	126	1
384	389	2	202	1 268	1 156	1 186	1 246	1 170	1 174	1 110	1 126	1 150	1 188
384			202	1 268	1 154	1 186	1 250	1 170	1 174	1 110	1 126	1 158	1 188
409	408	410	216	1 268	1 148	1 200	1 220	1 170	1 184	1 124	1 132	1 154	1 208
409			202	1 268	1 154	1 186	1 230	1 178	1 172	1 110	1 124	1 150	1 216
435	1	433	200	1 280	1 148	1 184	1 252	1 178	1 178	0 22	1 126	1 150	1 204
435			202	1 268	1 156	1 194	1 220	1 170	1 172	0 22	1 126	1 150	1 204
443	1	444	206	1 280	1 148	1 186	1 246	1 178	1 176	0 14	1 128	1 154	0 192
443			202	1 256	1 154	1 186	1 230	1 178	1 172	0 10	1 124	1 150	0 188
458	1	551	200	1 268	1 162	1 186	1 230	1 178	1 172	1 122	1 126	1 150	1 208
458			200	1 268	1 154	1 186	1 234	1 178	1 172	1 12	1 128	1 154	1 188
488	1	508	216	1 268	1 160	1 184	1 232	1 170	1 172	1 18	1 122	1 150	1 208
488			216	1 268	1 160	1 184	1 232	1 170	1 172	1 18	1 122	1 150	1 208
501	528	527	200	1 268	1 154	1 186	1 230	1 178	1 176	1 110	1 126	1 150	1 216
501			206	1 268	1 154	1 186	1 244	1 170	1 172	1 16	1 126	1 154	1 208
505	1	502	202	1 268	1 158	1 186	1 244	1 170	1 172	1 122	1 126	1 150	1 188
505			200	1 268	1 158	1 186	1 244	1 170	1 172	1 16	1 126	1 150	1 188
516	1	517	202	0 268	1 158	0	0	0	0	0 10	1 128	0	0 208
516			200	0 268	1 148	0	0	0	0	0 10	1 124	0	0 200
537	532	534	202	1 256	1 154	1 186	1 230	1 178	0 172	1 110	1 124	1 150	1 188
537			216	1 268	1 154	1 184	1 230	1 170	0 172	1 110	1 126	1 146	1 216
531	1	529	202	0 268	1 150	1 184	1 254	1 170	1 160	1 18	0 124	1 158	1 188
531			200	0 268	1 154	1 186	1 244	1 170	1 174	1 110	0 124	1 150	1 192
574	0	0	206	1 274	0 152	1 194	1 236	1 170	1 174	0 18	1 124	1 150	1 192
574			200	1 268	0 148	1 184	1 252	1 186	1 172	0 18	1 124	1 146	1 188
578	576	579	202	1 280	1 154	1 186	1 214	1 170	1 174	1 18	1 124	1 150	1 192
578			202	1 268	1 154	1 186	1 230	1 178	1 172	1 110	1 124	1 162	1 188

## Affected haplotypes

18p	IPAN	MAN	ca212	1140	59	ca231	1a201	at201	PDca225	w3442	ca213	ga201	ga203
587	580	582	202	1256	1158	1186	1248	1170	1174	116	1124	1150	1208
587		202		1268	1154	1186	1244	1170	1172	110	1132	1150	1208
361	1	360	204	0270	1158	1186	1244	1170	1172	110	1126	1150	1208
361		202		0276	1148	1186	1236	1170	1172	120	1128	1150	1212
368	0	0	204	1268	1164	1186	1242	0178	0172	110	1	0150	1192
368		202		1256	1154	1186	1230	0170	0160	110	126	0154	1212
374	1	2	200	1268	1154	1186	1230	1178	1174	010	1126	0150	0
374		200		1268	1154	1186	1230	1178	1160	010	1124	0142	0
399	0	0	202	1268	1148	1184	1	0170	1174	016	1124	1142	1188
399		204		1272	1158	1186	1	0178	1172	018	1126	1150	1200
411	1	2	216	0270	0164	0	184 0252	0	170 0174	018	0124	1150	0188
411		202		0268	0154	0	186 0230	0	178 0160	010	0124	1142	0188
413	414	412	200	1268	1158	1186	1230	1178	1178	118	1112	1150	1188
413		202		1280	1148	1186	1244	1170	1176	124	1126	1154	1188
236	697	698	216	1268	1158	1186	1220	1170	1172	120	1124	1150	1204
236		216		1268	1158	1186	1220	1170	1172	120	1124	1150	1204
421	0	0	200	1268	1148	1	0252	1170	1174	110	1126	1150	1188
421		202		1268	1152	1	0242	1190	1172	110	1126	1150	1188
424	1	2	200	1268	1158	0194	0	220 0	170 0178	024	0128	0150	0208
424		200		1268	1154	0186	0	232 0	178 0160	018	0112	0146	0192
452	1	2	202	0	256 0	148 0184	1252	0170	1174	016	0124	1158	0188
452		200		0	268 0	154 0184	1250	0170	1160	010	0124	1150	0188
473	1	472	202	1268	1162	1186	1246	1170	1180	122	0126	1150	1212
473		218		1268	1148	1186	1244	1170	1160	110	0124	1146	1188
484	482	2	200	1276	1148	1	0246	1170	1174	114	1124	1150	1188
484		206		1256	1154	1	0244	1170	1174	110	1126	1150	1212
487	1	486	200	1268	1158	1190	0248	1170	1174	112	0126	1158	1192
487		202		1278	1148	1186	0246	1182	1180	110	0112	1150	1208
331	1	476	202	0268	1154	1186	1234	1178	0172	124	1126	0158	0212
331		200		0268	1154	1186	1230	1170	0172	110	1112	0150	0188
489	0	0	202	1268	1158	1186	1244	1170	1172	110	1124	1150	1204
489		200		1268	1148	1202	1220	1178	1172	110	1132	1162	1208
498	1	635	200	1268	1160	1186	1246	1170	1172	114	1122	1150	1208
498		200		1268	1164	1186	1246	1170	1172	118	1112	1150	1188

## Affected haplotypes

18p	PAN	MAN	ca212	1140	59	ca231	1a201	at201	PDca225	w3442	ca213	ga201	ga203
566	0	0	1 268	1 148	1 148	1 202	1	220	1 174	1 10	1 124	1 150	1 212
566	1	2	1 268	1 154	1 154	1 186	1	230	1 172	1 10	1 128	1 150	1 208
514	1	2	0 268	1 154	1 154	1 186	0 230	1 178	1 172	0 10	1 124	1 154	0 192
514	1	633	0 268	1 154	1 154	1 184	0 230	1 178	1 168	0 10	1 124	1 146	0 192
536	1	633	1 270	0 148	0 148	1 184	1 254	1 170	1	168	1 132	0 162	1 212
536	1	2	1 268	0 154	0 154	1 186	1 252	1 178	1	172	1 124	0 154	1 188
605	1	2	0 268	1 158	1 158	0 198	0 244	0	170	1 16	0 124	1	200
605	1	2	0 268	1 150	1 150	0 186	0 220	0	178	1 10	0 124	1 150	188
540	539	562	1 268	1 154	1 154	1 186	1 230	1 178	1 172	1 10	1 124	1 150	1 216
540	1	216	1 268	1 148	1 148	1 186	1 230	1 194	1 172	1 22	1 112	1 154	1 212
684	1	730	0 268	1 158	1 158	1 186	1 232	1 178	1 160	1 24	1 112	1 150	1 212
684	1	200	0 268	1 154	1 154	1 186	1 244	1 170	1 160	1 10	1 126	1 150	1 188
608	1	2	0 268	1 156	1 156	0 192	0 244	0 170	1 178	0 22	0 126	1 150	1 204
608	1	202	0 268	1 154	1 154	0 186	0 220	0 170	1 174	0 10	0 126	1 150	1 188
637	1	638	1 268	1 162	1 162	0 186	1 250	1 182	1 172	1 10	1 124	1	142
637	1	200	1 268	1 154	1 154	0 186	1 230	1 178	1 172	1 10	1 124	1	150
649	647	646	1 268	1 154	1 154	1 186	1 230	1 178	1 172	1 10	1 124	1 150	1 188
649	1	200	1 270	1 162	1 162	1 184	1 250	1 170	1 180	1 10	1 112	1 154	1 188
653	1	652	1 280	0	0	0 184	1 230	1 178	1 184	1 20	1 128	1 154	1
653	1	200	1 268	0	0	0 186	1 230	1 178	1 168	1 22	1 112	1 150	1
491	1	2	0 268	1 158	1 158	0 194	0 256	0 178	0 180	0 22	0 124	1 158	0 204
491	1	202	0 268	1 148	1 148	0 184	0 230	0 170	0 174	0 10	0 124	1 154	0 188
493	1	2	0 282	0	0	0 186	1 242	1 170	1 174	0 16	0 124	1 158	0 212
493	1	200	0 268	0	0	0 186	1 242	1 170	1 172	0 14	0 124	1 150	0 204
506	1	2	0	0	0	0	0	0	0	0	0	0 150	1
506	1	2	0	0	0	0	0	0	0	0	0	0 150	1
661	660	662	1 278	1 156	1 156	1 198	1 220	1 170	1 174	1 20	1 126	1 154	1 204
661	1	200	1 268	1 148	1 148	1 184	1 250	1 186	1 174	1 18	1 120	1 150	1 188
667	666	668	1 268	1 154	1 154	1 186	1 214	1 170	1 160	1 22	1 124	1 146	1 212
667	1	202	1 268	1 162	1 162	1 186	1 246	1 178	1 172	1 18	1 112	1 158	1 188
669	670	671	1 268	1 162	1 162	1 186	1 258	1 186	1 174	1 18	1 126	1 150	1 188
669	1	200	1 268	1 154	1 154	1 186	1 244	1 170	1 160	1 10	1 126	1 150	1 188
676	1	678	0 268	1 158	1 158	1 190	1 244	1 178	1 172	1 16	1 126	1 158	1 188
676	1	200	0 280	1 148	1 148	1 184	1 252	1 178	1 172	1 22	1 126	1 150	1 216

## Affected haplotypes

18p	PAN	MAN	ca212	1140	59	ca231	la201	al201	PDca225	lw3442	ca213	ga201	ga203
681	1	2	202	0	256	0	1260	0	186	0	126	0	1192
681			200	0	268	0	1230	0	178	0	124	0	1188
351	354	2	202	1	268	1	1230	1	178	1	126	1	1188
351			216	1	268	1	1244	1	186	1	124	1	1208
355	1	2	216	0	272	0	1248	0	170	0	126	0	1188
355			204	0	268	0	1244	0	178	0	124	0	1188

Affected haplotypes

ca219	1105	ca209	ca202	1146	166d05	476
241	185	1173	1192	1272	1312	1271
233	199	1181	1196	1270	1304	1271
241	185	1173	1182	1274	0312	1273
245	1103	1177	1194	1270	0308	1267
241	185	1173	1198	0274	1308	0275
241	185	1173	1194	0274	1304	0271
241	187	1173	1182	1272	1300	1271
235	1101	1181	1196	1274	1312	1271
235	185	1173	1182	1274	1312	1271
243	185	1173	1192	1274	1316	1267
245	1103	1177	1194	0274	0312	1271
235	191	1181	1182	0270	0316	1271
241	185	1173	1182	1274	1312	1271
241	1103	1177	1196	1274	1312	1271
241	185	1173	1196	0270	1300	1271
235	185	1181	1190	0274	1312	1267
235	181	1173	1182	1274	1324	1271
223	183	1173	1192	1274	1300	1267
245	1103	1177	1196	1274	1312	1271
241	185	1173	1182	1270	1312	1265
241	1105	0173	1196	1270	1304	1267
241	1101	0173	1196	1270	1308	1271
241	187	0173	1192	1274	1312	1271
241	185	0173	1196	1274	1304	1267
245	197	1177	1194	1274	1312	0271
235	199	1181	1198	1274	1300	0271
241	095	0181	0198	0274	1320	0273
235	085	0173	0196	0274	1308	0271
235	1101	0181	0196	1272	1312	1271
235	185	0173	0200	1274	1308	1271
241	0	85	1200	0274	0312	0271
243	0	101	1196	0270	0304	0267
241	185	1173	1182	1274	1316	1271
241	199	1173	1200	1274	1300	1269

Affected haplotypes

ca219	1105	ca209	ca202	1146	166d05	476
241 185	1 177	1 196	1 270	1 304	1 271	1
245 199	1 177	1 192	1 274	1 308	1 265	1
245 197	1 177	1 196	0 274	1 304	1 275	0
245 199	1 177	1 192	0 270	1 308	1 267	0
243 1103	1 175	1 198	1 274	1 300	1 271	1
245 185	1 173	1 194	1 274	1 312	1 271	1
235 1101	0 181	0 196	1 270	1 316	1 267	1
235 185	0 173	0 196	1 274	1 300	1 271	1
241 185	1 173	1 200	0 270	1 304	1 273	1
241 185	1 177	1 198	0 274	1 308	1 271	1
241 0101	0 177	1 182	1 274	1 312	1 273	1
235 085	0 177	1 190	1 274	1 300	1 275	1
241 185	1 173	1 194	1 274	1 308	1 271	1
239 185	1 173	1 196	1 270	1 308	1 271	1
245 085	1 177	1 198	1 274	1 320	1 271	1
241 085	1 173	1 196	1 274	1 308	1 265	1
241 199	0 177	1 198	1 270	1 312	1 271	0
241 185	0 173	1 182	1 270	1 312	1 263	0
241 0101	0 187	0 200	0 270	1 312	0 271	1
235 085	0 173	0 182	0 270	1 300	0 271	1
241 0101	1 181	0 196	1 274	1 308	0 275	0
235 083	1 173	0 196	1 274	1 304	0 267	0
241 185	1 173	1 196	1 270	1 300	1 275	1
235 1101	1 181	1 196	1 272	1 300	1 271	1
241 185	1 173	1 182	1 270	1 300	1 271	1
239 1103	1 173	1 194	1 274	1 312	1 271	1
241 185	1 173	1 194	1 274	1 316	1 271	1
241 185	1 173	1 196	1 270	1 308	1 271	1
241 105	1 173	1 196	1 274	1 312	0 271	0
235 1101	1 181	1 182	1 270	1 300	0 267	0
243 185	1 173	1 196	1 274	1 300	1 271	1
241 185	1 173	1 190	1 270	1 316	1 271	1
241 185	1 177	1 196	1 274	1 304	1 271	0
241 185	1 173	1 196	1 270	1 312	1 267	0

Affected haplotypes

ca219	1105	ca209	ca202	1146	166d05	476
241	1	85 0 173	1 192	0 272	0 312	0 273
241	1	103 0 173	1 182	0 270	0 304	0 267
245	1 103	1 181	0 194	1 270	1 312	1 271
235	1 101	1 177	0 202	1 274	1 312	1 271
241	0 103	1 181	0 196	1 276	0 304	1 271
235	0 101	1 173	0 208	1 274	0 300	1 267
241	1 85	1 173	1 198	1 274	1 304	1 271
241	1 85	1 173	1 190	1 274	1 312	1 271
245	1 97	1 177	1 196	1 274	1 304	1 275
235	1 99	1 181	1 196	1 274	1 304	1 271
245	1 103	1 177	1 194	1 270	1 312	1 271
245	1 85	1 173	1 192	1 274	1 308	1 267
235	1 101	1 181	1 196	0 272	1 300	1 271
241	1 85	1 173	1 184	0 274	1 320	1 269
245	1 85	1 177	1 196	1 274	1 312	1 271
235	1 101	1 181	1 182	1 270	1 312	1 269
241	0 103	0 181	0 190	1 274	1 312	0 271
235	0 101	0 173	0 190	1 274	1 304	0 267
241	1 101	1 181	0 196	0 274	1 312	1 271
235	1 103	1 173	0 192	0 274	1 300	1 271
	0 101	1	0	0 270	1 304	1
	0 85	1	0	0 272	1 300	1
241	1 85	1 173	1 194	0 274	0 312	1 269
241	1 101	1 177	1 182	0 270	0 312	1 267
241	1 85	1 173	1 182	1 276	1 320	1 269
241	1 85	1 173	1 194	1 270	1 300	1 271
241	1 85	1 173	1 200	1 272	0 304	1 271
241	1 85	1 173	1 182	1 270	0 316	1 271
241	0 85	1 181	1 190	0 274	1 316	0 267
235	0 85	1 181	1 182	0 274	1 312	0 263
235	1 81	1 179	1 196	0 274	1 312	1 269
235	1 85	1 179	1 182	0 274	1 312	1 271
235	1	85 0 181	1 194	1 274	1 300	1 275
241	1	101 0 173	1 196	1 270	1 300	1 271

Affected haplotypes

ca219	1105	ca209	ca202	1146	166d05	476
243	0 103	0 177	0 196	0 274	0 308	1 271
241	0 85	0 173	0 190	0 270	0 312	1 265
235	1 99	1 181	1 196	1 274	1 308	1 271
235	1 101	1 181	1 196	1 272	1 308	1 267
241	1 85	1 177	0 192	1 270	1 316	0 269
245	1 85	1 173	0 184	1 274	1 308	0 265
241	1 99	1 177	1	0 274	0 308	1 267
241	1 105	1 173	1	0 270	0 300	1 271
241	0 103	1 181	0 190	1 274	1 312	1 271
235	0 97	1 173	0 198	1 270	1 300	1 267
241	1 99	1 177	0 182	1 274	1 308	1 271
241	1 85	1 173	0 196	1 274	1 300	1 271
245	1 85	1 177	1 182	1 274	1 312	1 273
245	1 85	1 177	1 182	1 274	1 312	1 267
241	1 85	1 175	1 196	1 274	1 320	1 261
241	1 101	1 173	1 196	1 270	1 304	1 267
241	0 85	1 173	1 186	1 270	1 316	1 269
239	0 85	1 173	1 182	1 270	1 312	1 273
235	1 101	1 181	1 184	1 274	1 324	1 269
235	1 101	1 181	1 184	1 274	1 324	1 269
241	1 85	1 173	1 190	1 274	0 316	1 271
245	1 101	1 175	1 196	1 270	0 308	1 271
241	1 85	1 173	1 196	0 270	1 316	1 267
243	1 85	1 173	1 192	0 274	1 308	1 267
241	0 99	0 181	0 196	0 274	1 312	1 271
235	0 85	0 173	0 192	0 274	1 312	1 267
241	1 101	0 173	1 196	1 270	1 304	1 267
241	1 85	0 173	1 194	1 270	1 312	1 267
241	1 99	1 173	1 192	1 274	1 312	1 271
225	1 83	1 173	1 192	1 270	1 308	1 269
241	1 85	1 173	1 182	1 274	0 312	0 271
241	1 85	1 181	1 182	1 270	0 308	0 269
245	1 103	1 177	1 196	0 270	1 304	1 267
241	1 105	1 173	1 192	0 274	1 316	1 271

Affected haplotypes

ca219	1105	ca209	ca202	1146	166d05	476
0 85	1 173	1 190	1 274	1 312	1 1271	1
0 101	1 181	1 198	1 272	1 312	1 1263	1
1 99	1 177	0 198	1 274	1 312	1 1271	1
1 101	1 173	0 196	1 276	1 304	1 1265	1
1 85	1 173	1 196	1 270	1 304	1 1271	1
1 85	1 173	1 190	1 270	1 312	1 1271	1
1 85	0 173	1 200	0 270	1 312	0 271	1
1 85	0 173	1 186	0 270	1 304	0 271	1
1 85	1 173	1 200	1 274	1 312	1 1271	1
1 95	1 181	1 196	1 274	1 312	1 1271	1
0 85	1 173	1 200	1 274	1 312	0 271	1
0 85	1 173	1 200	1 274	1 308	0 271	1
1 85	1 173	1 194	1 274	1 300	1 1275	1
1 85	1 181	1 196	1 274	1 300	1 1271	1
1 103	1 175	1 198	1 274	1 300	1 1271	1
1 103	1 175	1 196	1 274	1 308	1 1271	1
1 97	1 181	1 196	1 274	1 300	1 1271	1
1 99	1 181	1 192	1 270	1 312	1 1267	1
0 101	0 181	0 194	0 274	0 308	0 271	0
0 85	0 173	0 182	0 272	0 300	0 267	0
0 103	0 173	1 196	1 274	1 308	0 269	0
0 85	0 173	1 196	1 274	1 304	0 267	0
1 87	1 177	0 196	1 274	0 312	1 1271	1
1 85	1 173	0 194	1 270	0 300	1 1275	1
1 105	1 173	1 196	1 274	0 312	1 1271	0
1 85	1 173	1 192	1 270	0 312	1 1267	0
0 85	1 173	1 198	1 270	1 304	1 1271	1
0 85	1 173	1 196	1 274	1 312	1 1271	1
1 85	1 173	1 196	1 274	1 308	1 1271	0
1 85	1 173	1 182	1 274	1 320	1 1265	0
0 85	1 177	0 198	1 274	1 304	1 1271	0
0 85	1 173	0 194	1 274	1 312	1 1267	0
1 103	1 177	1 196	1 270	1 316	1 1267	1
1 99	1 181	1 192	1 270	1 312	1 1271	1

Affected haplotypes

ca219	1105	ca209	ca202	1146	166d05	476
245	1105	0177	1196	1274	1300	1267
245	185	0177	1198	1274	1320	1271
241	197	0177	0196	1274	0304	0271
241	185	0173	0196	1272	0300	0271
241	199	0177	0196	1274	1312	1271
241	185	0173	0182	1274	1312	1271
243	085	1173	1200	0274	1308	1271
235	085	1173	1194	0274	1308	1271
241	185	1173	1190	1274	0312	1267
235	185	1173	1196	1272	0316	1267
241	085	1181	0196	1274	0312	1269
235	0101	1173	0196	1272	0300	1271
245	0101	173	182	0274	1312	1273
241	085	177	190	0274	1312	1267
239	185	1173	1190	1270	1300	1271
241	185	1173	1198	1274	1304	1271
241	085	1173	1198	1270	1304	1271
243	085	1173	1182	1274	1312	1269
245	185	1179	1196	1270	1308	1271
241	185	1173	1196	1270	1304	1265
241	0103	0173	1198	0274	1308	1269
235	081	0173	1196	0274	1308	1265
241	1103	0177	0196	0270	1308	0271
241	185	0173	0190	0270	1300	0269
245	0	0	0	0	0	0
241	0	0	0	0	0	0
235	181	1173	1196	1276	1300	1271
241	185	1173	1196	1274	1308	1265
245	1103	1177	1196	1270	1308	1271
235	197	1181	1192	1274	1312	1271
241	1101	0173	1192	1274	1316	1271
235	185	0181	1190	1270	1312	1271
235	197	1181	1198	1274	1312	0271
243	1103	1173	1182	1274	1308	0273

Affected haplotypes

ca219	1105	ca209	ca202	1146	166d05	476
241	1 103	0 177	0 190	0 270	0 304	0 271
241	1 85	0 173	0 196	0 274	0 312	0 271
241	1 101	1 173	1 192	1 274	1 312	1 271
245	1 105	1 177	1 194	1 274	1 320	1 267
241	1 103	0 177	0 196	1 274	0 316	0 267
241	1 85	0 173	0 196	1 270	0 304	0 267

Fig 9

nontransmitted chromosomes

ERSN	KID	sava5	ca211	ca212	1140	59	ca231	ta201	ai201	ca225	w3442	ca213	ga201	ga203
279	200	235	1193	1216	1268	1148	1186	1246	1194	1172	1116	1124	1150	1188
280	200	233	1205	1202	1278	1148	1184	1252	1170	1172	120	1124	1150	1192
349	204	235	1197	1202	1268	1156	1184	1252	1170	1172	120	1120	1150	1216
309	204	235	1195	1202	1268	1148	1186	1244	1170	1172	116	1124	1142	1192
277	207	227	1205	1200	1268	1148	1184	1252	1186	1174	118	1124	1146	1212
278	207	227	1195	1200	1268	1158	1186	1230	1178	1168	120	1124	1150	1200
459	214	233	1197	1200	1268	1152	1184	1248	1186	1174	110	1124	1142	1208
460	214	233	1203	1216	1280	1158	1184	1248	1170	1184	116	1124	1146	1216
270	215	235	1193	1200	1268	1154	0188	1246	1170	1160	124	1124	1150	0188
259	216	231	1193	1200	1268	0150	1184	0254	1186	1172	110	1124	1150	0188
272	218	233	1195	1204	1268	1150	1186	1248	1178	1172	022	1126	1146	1188
273	218	235	1193	1200	1256	1154	1186	1230	1178	1172	010	1124	1142	1188
267	220	233	1205	1200	0268	1158	1186	1244	1170	1160	114	1124	1158	1188
264	225	227	1201	1200	1268	1150	1186	1242	1170	0168	010	0126	1150	1192
260	228	229	1197	1200	1268	1164	1186	1250	1178	1172	014	1112	1154	1188
257	229	227	1207	1200	1256	0160	1186	1246	1170	1172	114	1122	1150	1208
298	231	233	1193	1200	1280	1158	1186	1232	1178	1172	112	1112	1154	1188
299	231	229	1207	1200	1268	1148	1202	1220	1170	1160	114	1112	1158	1208
310	232	233	1205	1202	1268	1148	1204	1220	1170	1160	124	1112	1150	0188
261	234	233	1189	1206	1272	1154	1186	1250	1178	1174	118	1126	1158	1188
697	236	235	1197	1200	1268	1154	1186	1230	1186	1174	110	1112	1150	1208
698	236	233	1195	1202	1278	1148	1184	1252	1170	1172	120	1120	1150	1216
456	238	235	1199	1216	1268	1160	1184	1248	1170	1172	116	1124	1150	1208
457	238	233	1197	1200	1268	1160	1186	1230	1170	1172	118	1122	1150	1208
312	239	227	1197	1202	1268	1148	1184	0246	1170	1178	124	1112	1150	0208
342	241	227	1193	1202	1256	1158	1184	0250	1170	1174	110	1124	0146	1188
347	243	229	1193	0216	1278	1150	1186	1244	1170	0160	010	1112	1150	1188
274	243	233	1193	1204	1268	1160	1186	1244	1170	0160	014	1124	1162	1188
262	246	231	1193	0202	0268	1148	1202	1230	1170	0172	122	1124	1150	1208
302	247	235	1195	1200	1256	1150	1186	1242	1170	1172	110	1126	1150	1192
303	247	227	1195	1200	1268	1158	1186	1230	1178	1168	114	1128	1150	1188
334	248	225	1183	1216	1268	1152	1186	1230	1178	1176	110	1126	1150	1188
333	248	233	1205	1200	1268	1152	1186	1230	1178	1172	110	1124	1142	1188
300	251	227	0193	1200	1278	1148	1184	1252	1170	1172	118	1120	1150	1216

## nontransmitted chromosomes

ERSN	KID	sava5	ca211	ca212	1140	59	ca231	ta201	al201	ca225	w3442	ca213	ga201	ga203
301	251	227	0 205	1 200	1 276	1 148	1 184	1 252	1 170	1 172	1 24	1 124	1 150	1 188
285	252	231	1 193	1 200	0 268	1 148	1 184	1 252	1 170	1 174	1 16	1 124	0 150	1 192
258	253	229	1 193	1 200	0 268	1 148	1 186	1 230	1 194	1 172	1 22	1 112	1 154	1 208
467	254	229	1 197	1 216	1 280	1 160	1 184	1 250	1 170	0 172	1 22	1 126	1 154	1 188
266	265	227	0 195	1 202	1 268	1 160	1 186	1 260	1 178	1 174	1 16	1 124	0 158	1 208
485	311	227	1 205	1 200	1 268	1 158	1 184	1 230	1 178	1 184	1 20	1 128	1 154	1 212
313	314	227	1 195	1 202	1 268	1 162	1 186	1	0 170	1 172	1 10	0 124	1 150	1 212
348	314	227	1 195	1 200	1 268	1 148	1 184	1 248	1 170	1 172	1 10	0 128	1 150	1 208
317	316	227	1 201	1 202	1 268	1 152	1 186	1 244	1 170	1 174	1 14	1 112	1 154	1 188
318	319	227	0	0	0 256	0 154	0	0	0	0	0 16	1	0	0
320	321	237	1 201	0 200	0 268	1 154	0 186	0 220	1 170	1 172	0 20	1 124	0 146	0 192
336	326	227	1 193	1 202	1 268	1 154	1 186	1 244	1 170	1 160	1 18	1 124	1 154	1 208
325	326	227	1 201	1 202	1 276	1 148	1 186	1 244	1 170	1 176	1 20	1 126	1 150	1 192
330	329	233	1 197	1 202	1 268	1 148	0 184	1 256	1 178	1 172	1 16	1 124	1 162	1 208
476	331	229	0 199	1 200	0 276	1 154	1	0 244	1 170	0 160	1 10	1 112	0 150	0 188
354	351	233	1 201	0 200	1 268	1 162	1 186	1 248	1 178	1 160	1 22	1 132	1 150	1 188
352	353	225	0 207	1 200	1 268	0 154	1 194	1 220	1 170	1 178	1 18	1 128	1 146	1 192
362	356	231	1 195	1 202	1 268	1 154	1 186	0 230	1 170	0 172	1 10	0 128	1 150	1 208
358	357	235	1 205	1 202	0 256	1 154	0 186	1 230	1 178	1 172	1 10	1 124	0 154	1 216
365	359	233	1 205	1 200	1 268	1 162	1 186	1 248	1 178	1 160	1 22	1 132	1 150	1 188
378	359	231	1 201	1 202	1 268	1 162	1 186	1 230	1 186	1 174	1 18	1 126	1 150	1 188
360	361	227	0 195	1 202	0 268	1 162	1 186	1 250	1 170	1 172	1 18	1 124	1 150	1 212
366	367	227	1 193	1 202	1 268	1 154	1 186	1 230	1 178	1 160	1 10	1 124	1 142	0 188
370	372	227	0 201	1 202	1 268	1 150	1 184	0 244	0 170	1 174	1 14	1 124	0 150	1 188
389	384	231	1 203	1 204	1 272	1 158	1 186	1 244	1 178	1 172	1 18	1 126	1 150	1 200
408	409	229	1 205	1 216	1 276	1 154	1 186	1 244	1 178	1 184	1 28	1 112	1 154	1 196
410	409	229	1 197	1 204	1 272	1 158	1 186	1 244	1 178	1 172	1 18	1 126	1 150	1 188
414	413	227	1 195	1 200	1 268	1 158	1 186	1 242	1 178	0 174	1 18	1 120	1 150	1 188
412	413	235	1 193	1 200	1 256	1 156	1 186	1 246	1 170	1 172	1 10	1 124	1 150	1 212
433	435	227	1 195	1 202	1 268	1 154	1 186	1 242	1 170	1 172	0 16	1 112	1 150	1 204
444	443	235	1 205	1 200	1 268	1 158	1 186	1 232	1 178	1	0 24	1 112	1 150	0 188
551	458	235	1 201	1 206	1 268	1 148	1 184	1 248	1 170	1 174	1 14	1 124	1 158	1 188
472	473	233	1 193	1 200	1 268	1 156	1 186	1 248	1 178	1 184	1 10	0 112	1 146	1 188
482	484	233	0 197	1 200	1 268	1 158	1 182	0 248	1 170	1 174	1 16	1 124	1 150	1 188

## nontransmitted chromosomes

ERSM	KID	sava5	ca211	ca212	1140	59	ca231	ta201	at201	ca225	w3442	ca213	ga201	ga203
486	487	227	1201	1202	1256	1154	1186	0230	1178	1172	110	0124	1150	1188
508	488	233	1205	1202	1268	1148	1184	1220	1170	1160	124	1112	1150	1188
635	498	227	1193	1202	1268	1148	1184	1254	1170	1174	116	1124	1142	1188
527	501	229	1183	1216	1280	1158	1186	1230	1170	1172	122	1126	1146	1212
528	501	225	1183	1216	1268	1152	1186	1242	1170	1176	110	1126	1154	1208
502	505	235	1205	1200	1268	1148	1184	1	0170	1174	110	1126	1150	1188
517	516		0	0	0	0	0	0	0	0	0	0	0	0
529	531	233	1205	1200	0268	1158	1186	1242	1170	1180	110	0128	1150	1204
633	536	229	0201	1200	1268	0154	1186	1230	1178	1168	010	1124	0150	1
532	537	227	1201	1200	1268	1150	1186	1242	1170	0172	110	1126	1150	1192
534	537	235	1205	1200	1268	1158	1186	1232	1170	0160	124	1112	1150	1208
562	540	229	1195	1202	1268	1160	1184	1250	1170	1160	118	1124	1150	1212
539	540	229	1207	1200	1268	1154	1194	1220	1170	1178	118	1128	1150	1192
576	578	235	1199	1200	1256	1158	1186	1246	1170	1174	110	1124	1158	1188
579	578	233	1199	1200	1278	1148	1186	1246	1170	1184	116	1124	1150	1208
582	587	227	1201	1202	1268	1148	1202	1220	1178	1184	110	1128	1150	1212
580	587	229	1	0200	1268	1154	1186	1244	1170	1160	110	1126	1150	1200
638	637	237	1203	1206	1268	1154	0186	1228	1170	1160	122	1126	1142	0212
647	649	229	1195	1202	1268	1154	1186	1232	1178	1160	110	1124	1150	1216
646	649	231	1201	1206	1268	1154	1186	1230	1178	1160	110	1124	1154	1188
652	653	235	1201	1206	1268	0154	0186	1230	1178	1172	116	1126	1150	1188
662	661	235	1209	1202	1280	1154	1186	1242	1178	1172	122	1126	1150	1188
660	661	233	1183	1216	1268	1158	1186	1	0170	1160	114	1122	1150	1188
666	667	235	1203	1202	1268	1158	1186	1246	1170	1174	110	1126	1150	1192
668	667	237	1209	1202	1268	1150	1186	1252	1178	1172	116	1128	1150	1196
670	669	235	1205	1200	1268	1148	1184	1254	1170	1174	110	1126	1154	1192
671	669	227	1195	1200	1268	1158	1186	1230	1178	1168	116	1128	1154	1188
678	676	223	1201	1200	0278	1156	1200	1252	1174	1174	110	1124	1150	1208
730	684	229	1195	1200	0268	1148	1198	1220	1170	1174	120	1126	1150	1196

nontransmitted chromosomes

ca219	1105	ca209	ca202	1146	166d05	476	
241	1 103	1 173	1 186	1 274	1 316	1 269	1
241	1 85	1 173	1 182	1 270	1 316	1 263	1
243	1 85	1 177	1 192	1 270	0 312	1 265	1
241	1 85	1 173	1 192	1 270	0 312	1 267	1
241	1 85	1 173	1 198	1 274	1 308	1 271	1
245	1 101	1 175	1 196	1 274	1 316	1 267	1
245	1 101	1 177	1 190	1 274	1 312	1 267	1
241	1 85	1 173	1 202	1 270	1 312	1 269	1
235	1 95	1 181	1 190	0 274	1 308	1 267	0
241	1 85	1 173	1 196	1 274	1 304	1 267	0
235	1 103	1 181	1 196	1 274	1 312	1 265	1
235	1 99	1 181	1 196	1 274	1 304	1 271	1
241	1 85	0 173	1 192	1 270	1 312	1 271	1
241	1 85	0 173	0 196	1 270	1 304	1 271	1
235	1 93	1 181	1 196	1 274	1 308	1 269	0
241	1 103	1 177	1 196	1 270	1 316	1 267	1
235	1 97	1 181	1 198	1 274	1 300	1 271	1
245	1 85	1 177	1 192	1 270	1 300	1 271	1
241	1 85	0 173	0 196	1 274	1 308	1 271	1
235	1 95	1 181	1 198	0 274	1 300	1 267	1
241	1 85	1 173	1 196	1 274	1 300	1 267	1
243	1 85	1 177	1 192	1 270	1 312	1 265	1
241	0 99	1 177	1 198	1 270	1 312	1 263	1
241	0 97	1 177	1 196	1 274	1 304	1 275	1
245	1 85	0 177	1 196	1 272	1 308	1 263	0
235	0 99	1 173	0 196	1	0 304	0	0
235	1 101	1 181	1 194	1 274	1 308	1 267	1
241	1 85	1 177	1 195	1 274	1 304	1 271	1
245	1 85	1 177	1 198	1 270	1 300	0 267	0
241	1 85	1 173	1 196	1 270	1 304	1 271	1
239	1 85	1 181	1 196	1 276	1 300	1 267	1
241	1 85	1 181	1 194	1 274	1 324	1 267	0
241	1 99	1 181	1 196	1 274	1 304	1 267	0
243	1 85	1 177	0 192	1 270	1 312	1 265	1

nontransmitted chromosomes

ca219	1105	ca209	ca202	1146	166d05	476	
235	1 101	1 177	0 200	1 272	1 316	1 267	1
235	0 85	1 173	0 192	1 274	0 308	1 267	1
235	1 101	1 181	1 196	1 274	1 308	1 265	1
245	1 103	1 175	1 198	1 274	1 300	1 271	0
235	1 101	1 181	1 202	1 274	1 316	1 265	1
245	1 85	1 179	1 184	0 270	1 308	1 269	0
241	1 85	1 173	1 192	1 270	1 312	1 269	1
241	1 85	1 173	1 198	1 270	1 308	1 271	1
235	0 101	0 173	0 190	1 274	1 304	0 267	0
245	1 85	1	0	0 274	1 320	1 269	1
	0 103	1 173	1 182	1 274	1 312	1 271	0
241	1 85	1 173	1 182	1 270	1 312	1 273	1
241	1 85	1 177	1 200	1 274	1 308	1 263	1
235	1 85	1 173	1 196	1 270	0 316	1 265	1
241	1 85	1 173	1 182	1 270	1 300	1 265	0
241	1 85	1 173	1 182	1 270	1 308	1 267	0
241	1 87	1 173	1 182	0 272	1 300	1 271	1
245	1 85	0 177	1 198	1 274	1 300	1 271	0
241	0 85	0 173	0 190	0 270	0 312	1 273	1
241	1 85	1 173	1 182	1 270	1 312	1 267	0
241	1 85	1 177	1 192	1 270	1 312	1 267	0
241	1 85	1 173	0 192	1 270	1 308	1 269	1
243	1 85	1 173	0 200	1 274	1 308	0 265	0
243	1 85	1 173	1 190	0 270	0 316	1 273	1
235	0 95	1 173	0 196	1 274	1 312	1 271	1
243	1 85	1 173	0 198	1 270	1 300	1 271	1
241	1 85	1 173	0 196	1 274	1 316	1 267	1
241	1 85	1 173	1 194	1 270	1 316	1 265	1
241	1 85	1 173	1 200	1 274	1 316	1 271	1
241	1 85	1 173	1 194	1 270	1 300	1 271	1
235	1 105	1 181	1 200	1 272	1 316	1 267	1
239	0 101	1 173	1 196	1 274	1 300	1 271	1
241	1 85	1 173	0 192	1 270	0 316	1 265	1
241	1 83	1 173	1 196	1 270	0 304	1 267	0

nontransmitted chromosomes

ca219	1105	ca209	ca202	1146	166d05	476	
241	0 103	1 173	1 192	1 274	1 312	1 267	1
243	1 85	1 173	1 196	1 274	1 308	1 273	1
243	1 85	1 173	1 200	1 274	1 312	1 271	1
241	1 85	1 173	1 182	1 270	0 320	1 267	1
241	1 87	1 173	1 198	1 270	0 312	1 267	1
235	1 97	1 181	1 192	0 274	1 300	1 271	1
	0	0	0	0 272	1	0	0
235	1 81	1 173	1 182	1 278	1 320	1 261	1
241	1 85	0 173	0 200	1 270	1 304	1 271	1
241	1 85	0 173	1 196	1 270	1 304	1 271	1
235	1 85	0 181	1 194	1 274	1 308	1 267	1
239	1 85	1 173	1 194	1 272	0 316	1 271	1
241	1 85	1 173	1 182	1 272	0 300	1 271	1
241	1 105	1 173	1 192	0 274	1 312	1 267	1
241	1 87	1 173	1 192	0 272	1 304	1 275	1
	0 103	1 173	1 194	1 270	1 316	1 271	1
	0 101	1 173	1 196	1 272	1 308	1 271	1
241	1 87	1 173	1 182	1 274	1 320	1 269	1
241	0 85	1 173	1 194	1 270	1 312	1 267	1
241	0 85	1 173	1 196	1 274	1 300	1 271	1
235	1 99	1 181	1 192	1 274	1 312	1 267	1
235	1 101	1 181	1 196	1 272	1 300	1 271	1
235	1 85	1 179	1 196	1 274	1 312	1 271	1
241	1 85	1 173	1 192	1 270	1 312	1 271	1
241	1 87	1 173	1 182	1 270	1 316	1 273	1
235	1	0 181	1 196	1 274	1 300	1 271	1
239	1 85	0 181	1 196	1 276	1 300	1 267	1
241	1 83	1 177	1 182	1 276	1 308	0 269	1
235	0 93	1 173	0 202	1 272	0 300	1 273	1
			79				

FG 10

controls

cont	sav5	ca211	ca212	1140	59	ca231	la201	al201	ca225	w3442	ca213	ga201	ga203
98	miss	193	1200	1	0156	1186	1230	1178	1176	110	1126	1150	1208
98	17	193	1216	1	0148	1186	1244	1178	1172	118	1124	1150	1208
99	193	193	1206	1268	1150	1184	1252	1178	1172	120	1124	1142	1204
99	195	195	1200	1268	1154	1184	1220	1170	1170	110	1128	1150	1188
101		189	1206	1272	1154	1186	1260	1178	1174	120	1126	1158	1216
101		203	1200	1268	1150	1186	1244	1170	1160	114	1122	1150	1188
102		195	1202	1268	1150	1202	1220	1178	1172	124	1124	1150	1212
102		205	1200	1268	1162	1186	1248	1178	1160	122	1132	1150	1188
104		195	1200	1268	1154	1186	1244	1170	1160	110	1126	1150	1188
104		203	1216	1268	1156	1186	1244	1186	1174	114	1126	1150	1192
105		193	1202	1268	1156	1186	1244	1170	1172	110	1126	1150	1188
105		201	1216	1268	1148	1186	1246	1194	1172	116	1124	1150	1188
107			0206	1268	1154	1186	1246	1170	1176	122	0	0154	1188
107			0202	1274	1150	1184	1246	1170	1174	116	0	0150	1216
108		201	0200	1268	1162	1186	1230	1178	1172	122	1126	1150	1188
108		195	0202	1280	1154	1186	1242	1178	1172	122	1126	1150	1192
110		199	1218	1268	1160	1184	1248	1170	1172	116	0124	1150	1208
110		205	1200	1268	1148	1184	1254	1170	1174	110	0126	1150	1188
111		193	1202	1268	1154	1186	1232	1178	1160	1	0124	1150	1188
111		191	1202	1268	1150	1184	1252	1170	1160	1	0128	1150	1188
114		207	1202	1268	1150	1200	1220	1170	1174	024	1126	1150	1212
114		195	1200	1278	1154	1186	1252	1178	1172	018	1124	1150	1192
113		191	1202	1276	1150	1184	1250	1170	1174	022	1124	1146	1216
113		207	1216	1268	1150	1186	1244	1170	1172	016	1124	1150	1192
116		193	1202	1268	1154	1186	1230	1178	1172	110	1	0124	1188
116		195	1202	1268	1154	1186	1248	1170	1172	110	1	0126	1212
117		201	1200	1268	1154	1186	1232	1178	1172	110	1	0124	1212
117		195	1202	1268	1160	1186	1256	1178	1174	116	1	0126	1212
119		193	1200	1270	1162	1186	1244	1170	1172	118	0124	1150	1216
119		193	1206	1268	1154	1186	1230	1178	1172	110	0	0126	1188
120		193	1216	1276	1158	1186	1242	1178	1174	118	0	0112	1192
120		203	1204	1272	1158	1186	1244	1178	1172	110	0	0126	1200
122		183	1200	1268	1154	1186	1242	1178	1160	116	1	0124	1204
122		195	1218	1268	1156	1186	1232	1178	1160	126	1	0124	1188

## controls

cont	sava5	ca211	ca212	1140	59	ca231	la201	at201	ca225	w3442	ca213	ga201	ga203
123		193	1200	1268	1150	1184	1252	1170	1160	110	1126	1154	1188
123		195	1216	1268	1154	1184	1232	1170	1160	120	1112	1150	0192
125		203	1200	1268	1148	1184	1252	1	0174	118	1124	1	0212
125		205	1202	1268	1148	1188	1250	1	0172	116	1124	1	0192
126		205	1200	1268	1148	1186	1248	1170	1160	114	1128	0150	1188
126		195	1204	1268	1150	1186	1246	1178	1172	122	1126	0150	1208
128		193	1200	1256	1158	1186	1	0170	1174	114	1112	1158	1188
128		191	1200	1268	1160	1184	1	0170	1172	118	1122	1150	1208
129		193	1206	1256	1154	1186	1244	1170	1174	110	1112	1158	1188
129		195	1216	1268	1150	1184	1250	1170	1172	114	1126	1150	1192
131		201	0200	1268	1154	1186	0252	0186	0176	018	1126	1150	1188
131		197	0200	1268	1150	1184	0244	0170	0172	010	1126	1150	1188
132		205	0200	1268	1148	1186	0252	0186	0174	018	1124	1150	1212
132		203	0200	1268	1158	1184	0248	0170	0172	018	1124	1158	1208
134		193	1216	1268	1148	1186	1220	1170	1174	114	1124	1150	1208
134		205	1202	1266	1160	1186	1230	1194	1172	122	1112	1154	1208
135		193	1202	1268	1154	1186	1244	1170	1160	118	1124	1154	1208
135		205	1202	1268	1154	1184	1230	1178	1184	120	1128	1154	1208
138		193	1202	1268	1154	1186	1230	1178	1	010	1124	1	0216
138		207	1200	1280	1148	1184	1252	1178	1	020	1126	1	0216
137		193	1206	1268	1154	1186	1230	1178	1	010	1126	1150	1192
137		201	1216	1270	1148	1184	1256	1186	1	010	1126	1150	1212
144			0200	1256	1154	1186	1	0	0174	110	1126	1150	1208
144			0206	1268	1154	1186	1	0	0176	122	1124	1150	1188
68		195	1202	1268	1164	1186	1	0	0172	122	1126	1150	1208
68		193	1202	1268	1160	1186	1	0	0172	118	1122	1150	1208
69		195	1218	1268	1148	1186	1246	1	0160	110	1124	1146	1208
69		201	1216	1268	1158	1186	1230	1	0172	120	1124	1150	1204
72		193	1200	1268	1148	1184	1	0170	1174	116	1124	1150	1188
72		193	1206	1256	1156	1186	1	0170	1172	110	1124	1150	1192
71		193	1216	1268	1146	1192	1248	1170	1174	116	1124	1154	1196
71		193	1206	1256	1156	1186	1232	1170	1174	110	1126	1150	1212
74		195	1218	1268	1148	1186	1246	1170	1160	110	1124	1154	1216
74		205	1200	1268	1158	1186	1222	1170	1160	124	1112	1154	1188

controls

cont	sava5	ca211	ca212	1140	59	ca231	la201	ai201	ca225	w3442	ca213	ga201	ga203
75		217	1 216	1 264	1 150	1 186	1 250	1 170	1 180	1 12	1 124	1 150	1 192
75		205	1 204	1 268	1 154	1 186	1 244	1 170	1 172	1 16	1 124	1 146	1 192
78		201	1 216	1 268	1 148	1 186	1	0 174	1 172	1	0 124	1 150	1 192
78		201	1 202	1 268	1 162	1 186	1	0 170	1 174	1	0 126	1 150	1 188
77		201	1 206	1 268	1 158	1 184	1 246	1 170	1 160	1 22	1 124	1 150	1 192
77		195	1 202	1 268	1 152	1 186	1 232	1 178	1 174	1 20	1 122	1 146	1 188
80		193	1 202	0 268	1	0 186	1 250	1 178	1 160	1 10	1 124	1 150	1 208
80		195	1 200	0 268	1	0 186	1 244	1 178	1 172	1 28	1 124	1 150	1 208
81		193	1 202	0 268	1 156	1 186	1 246	1 194	1 172	1 10	1 126	1 150	1 188
81		193	1 200	0 268	1 148	1 184	1 258	1 186	1 174	1 10	1 124	1 150	1 208
84		193	1 202	1 268	1 154	1 186	1 246	1 170	1 172	1 14	1 126	1 158	1 188
84		207	1 202	1 268	1 164	1 186	1 244	1 170	1 178	1 10	1 124	1 150	1 188
83		209	1 200	1 270	1 148	1 184	1 230	1 178	1 172	1 26	1 124	1 150	1 208
83		207	1 200	1 268	1 158	1 186	1 248	1 170	1 174	1 10	1 112	1 146	1 192
86		195	1 202	1 268	1 158	1 186	1 244	1 170	1 160	1 14	1 124	1 158	1 208
86		205	1 202	1 278	1 148	1 184	1 260	1 170	1 172	1 20	1 120	1 150	1 188
87		197	1 200	1 268	1 158	1 186	1 230	1 178	1 172	1 10	1 124	1 158	1 188
87		193	1 200	1 268	1 154	1 190	1 242	1 170	1 172	1 16	1 126	1 154	1 188
90		205	1 200	1 268	1 158	1 186	1 250	1 170	1 172	1 18	1 124	1 158	1 208
90		193	1 200	1 268	1 154	1 186	1 246	1 186	1 172	1 10	1 124	1 150	1 188
89		207	1 202	1 270	1 168	1 186	1 232	1 178	0 176	1 22	1 126	1 154	1 212
89		193	1 202	1 268	1 154	1 190	1 252	1 170	0 172	1 16	1 126	1 150	1 188
92		193	1 200	1 268	1 148	1	0 244	1	0 174	1 10	1 124	1 150	1 208
92		193	1 202	1 256	1 154	1 186	0 230	1 178	0 172	1 10	1 124	1 154	1 188
93		203	1 216	1 268	1 156	1 186	0 248	1 170	1 174	1 14	1 126	1 150	1 204
93		205	1 200	1 268	1 148	1 184	0 230	1 178	1 174	1 10	1 126	1 150	1 188
95		197	1 216	1 268	1 158	1 186	1 252	1 178	1 174	1 20	1 126	1 150	1 192
95		205	1 202	1 268	1 150	1 184	1 230	1 178	1 160	1 10	1 126	1 150	1 188
96		209	1 200	1 278	1 162	1 186	1 256	1 170	1 160	1 14	1 128	1 150	1 192
96		205	1 200	1 268	1 148	1 186	1 230	1 178	1 160	1 14	1 128	1 150	1 188
140			0	0 270	1	0	0 244	1	0	0 10	1	0 150	1 188
140			0	0 278	1	0	0 254	1 186	1	0 10	1	0 158	1 188
141		201	0 200	1 272	1	0	0 244	1 170	1 172	1 10	1	0	0 216
141		193	0 200	1 270	1	0	0 254	1 170	1 160	1 10	1	0	0 212

## controls

cont	sava5	ca211	ca212	1140	59	ca231	la201	al201	ca225	w3442	ca213	ga201	ga203
143		193	1200	1278	148	184	1252	170	184	118	124	150	192
143		195	1200	1268	158	186	1248	1	178	118	124	146	188

controls

ca219	1105	18SCA20	SC	KID
241	1	missing	173	1 100
241	1	177	177	1 100
235	1	173	173	1 100
245	1	175	175	1 100
235	1	173	173	1 103
235	1	181	181	1 103
241	1	173	173	1 103
241	1	173	173	1 103
235	1	181	181	1 106
241	1	173	173	1 106
241	1	173	173	1 106
241	1	173	173	1 106
241	1	173	173	1 109
241	1	173	173	1 109
241	1	173	173	1 109
245	1	175	175	1 109
235	1			0 112
235	1			0 112
243	1	181	181	0 112
235	1	173	173	0 112
241	1	173	173	1 115
241	1	173	173	1 115
241	1	173	173	1 115
241	1	173	173	1 115
241	1	173	173	1 115
241	1	173	173	1 118
241	1	173	173	1 118
241	1	173	173	1 118
241	1	177	177	1 118
241	1	173	173	1 121
241	1	173	173	1 121
241	1	173	173	1 121
241	1	181	181	1 121
235	0	173	173	1 124
	0	173	173	1 124

controls

ca219	1105	18SCA20	SC	KID
241	1	173	1	124
241	1	177	1	124
241	1	173	1	127
241	1	173	1	127
243	1	173	1	127
239	1	173	1	127
235	1	181	1	130
235	1	181	1	130
241	1	173	1	130
241	1	173	1	130
243	0	181	0	133
235	0	173	0	133
245	0	181	0	133
235	0	173	0	133
243	1	173	1	136
235	1	181	1	136
241	1	173	1	136
241	1	173	1	136
241	1		0	139
243	1		0	139
235	1	181	0	139
241	1	177	0	139
241	1	173	1	145
241	1	173	1	145
241	1	173	1	170
245	1	177	1	170
241	1	173	1	170
243	1	175	1	170
243	1	173	1	173
241	1	173	1	173
235	1	181	1	173
241	1	173	1	173
241	1	173	1	176
235	1	181	1	176

controls

ca219	1105	18CA20	SC	KID
241	1	175	1	176
235	1	177	1	176
241	1	173	1	179
241	1	177	1	179
241	1	173	1	179
241	1	181	1	179
241	1		0	82
241	1		0	82
241	1	173	1	182
241	1	173	1	182
241	1	173	1	185
241	1	173	1	185
241	1	173	1	185
241	1	173	1	185
241	1	173	1	188
	0	173	1	188
	0	173	1	188
235	1	173	1	188
235	1	181	1	188
245	1	173	1	191
241	1	173	1	191
241	1	181	1	191
241	1	173	1	191
241	1	173	1	194
241	1	173	1	194
241	1	177	1	194
235	1	181	1	194
241	1	173	1	197
245	1	173	1	197
221	1	173	1	197
241	1	173	1	197
	0		0	142
	0		0	142
241	1	173	1	142
241	1	173	1	142

controls

ca219	1105	18SCA20	SC	KID
241	1	177	1	145
235	1	181	1	145

FIG 11

## AHR RESULTS IN DISEASE CHROMOSOMES

	SAVAS	CA211	CA212	18S1140	18S59	TA201	CA231	AT201	CA225	W344
SAVAS	0.04 2%				0.07 4%					
CA211	2-3									
CA212										
18S1140					0.70 12%		1.92 30%		1.28 20%	
18S59	2-6			2-4		2.11 10%	2.45 18%	0.53 10%	1.16 12%	4.18 17%
TA201					4-3		0.68 3%	1.34 10%	0.02 2%	0.68 8%
CA231				2-2	4-2	2-4			0.89 16%	
AT201					4-2	3-2			0.18 4%	0.03 4%
CA225				2-3	4-3	3-3	2-3	1-1		
W3442					4-1	3-1	2-1	2-1		

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/14892

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12Q 1/68; C12P 19/34; C07H 21/02, 21/04

US CL : 435/6; 91.2; 536/23.1, 24.3, 24.33

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6; 91.2; 536/23.1, 24.3, 24.33

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Coon et al. Analysis of Chromosome 18 DNA Markers in Multiplex Pedigrees with Manic Depression. Biological Psychiatry. 15 April 1996. Volume 39, pages 689-696, see entire document.	1-16
A	Stine et al. Evidence for Linkage of Bipolar Disorder to Chromosome 18 with a Parent-of-Origin Effect. American Journal of Human Genetics. December 1995. Volume 57, pages 1384-1394, see entire document.	1-16

☐

Further documents are listed in the continuation of Box C.

☐

See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*B* earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A* document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means	
*P* document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search 28 SEPTEMBER 1997	Date of mailing of the international search report 17 DEC 1997
Name and mailing address of the ISA/US Commissioner of Patents and Trademarks Box PCT Washington, D.C. 20231 Facsimile No. (703) 305-3230	Authorized officer DIANNE REES Telephone No. (703) 308-0196

Form PCT/ISA/210 (second sheet)(July 1992)\*

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/14892

## B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, BIOSIS, BIOTECHABS, BIOTECHDS, CABA, CAPLUS, DISSABS, DRUGU, EMBASE, EUROPATFULL, EUROPEX, MEDLINE, SCISEARCH, WPIDS, TOXLINE, TOXLIT, USPATFULL

search terms: bipolar, manic depression, mood disorder, shizoaffective disorder, chromosome 18, 18p, short arm, and markers, D18S1140, ga203, SAVA5, W3422, a201, ta201.



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

<p>(51) International Patent Classification<sup>6</sup> : <b>C12Q 1/68, C12P 19/34, C07H 21/02, 21/04</b></p>	<p><b>A1</b></p>	<p>(11) International Publication Number: <b>WO 98/07887</b> (43) International Publication Date: 26 February 1998 (26.02.98)</p>
<p>(21) International Application Number: PCT/US97/14892 (22) International Filing Date: 22 August 1997 (22.08.97) (30) Priority Data: 60/023,438 23 August 1996 (23.08.96) US (71) Applicant: THE REGENTS OF THE UNIVERSITY OF CALIFORNIA [US/US]; 22nd floor, 300 Lakeside Drive, Oakland, CA 94612-3550 (US). (72) Inventors: FREIMER, Nelson, B.; 630 29th Street, San Francisco, CA 94131 (US). LEON, Pedro; Centro de Investigaciones Biologia, University of Costa Rica, P.O. Box 2060, San Jose (CR). REUS, Victor, I.; 1214 Third Avenue, San Francisco, CA 94122 (US). SANDKUJIL, Lodewijk, A.; Voorstraat 27A, NL-2611 JK Delft (NL). McINNES, Lynne, Allison; 1599 Shrader Street, San Francisco, CA 94117 (US). SERVICE, Susan, K.; 816 Maher Road, Watsonville, CA 95076 (US). (74) Agents: HUGHES, Melya, J.; Cooley Godward LLP, 3000 El Camino Real, Five Palo Alto Square, Palo Alto, CA 94306-2155 (US) et al.</p>		<p>(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, UZ, VN, YU, ZW, ARIPO patent (GH, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).</p> <p><b>Published</b> <i>With international search report. Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i></p>
<p>(54) Title: METHODS FOR TREATING BIPOLAR MOOD DISORDER ASSOCIATED WITH MARKERS ON CHROMOSOMES 18p</p>		
<p>(57) Abstract</p> <p>The present invention is directed to methods of detecting the presence of a bipolar mood disorder susceptibility locus in an individual, comprising analyzing a sample of DNA for the presence of a DNA polymorphism on the short arm of chromosome 18 between the telomere and D18S481, wherein the DNA polymorphism is associated with a form of bipolar mood disorder. The invention for the first time provides strong evidence of a susceptibility gene for bipolar mood disorder that is located in the terminal 5 cM region of the short arm of chromosome 18. The disclosure describes the use of linkage analysis and genetic markers in this 5 cM region to fine map the region and the use of genetic markers to genetically diagnose (genotype) bipolar mood disorder in individuals, to confirm phenotypic diagnoses of bipolar mood disorder, to determine appropriate treatments for patients with particular genotypic subtypes. Isolated polynucleotides useful for genetic linkage analysis of BP-I and methods for obtaining such isolated polynucleotides are also described.</p>		

**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

# METHODS FOR TREATING BIPOLAR MOOD DISORDER ASSOCIATED WITH MARKERS ON CHROMOSOME 18p

## 5 . ACKNOWLEDGEMENTS

This invention was made with Government support under Grant Nos. RO1-MH49499, K21MH00916, awarded by the NIH. The U.S. Government has certain rights in this invention.

## 10 INTRODUCTION

## Background

## BIPOLAR MOOD DISORDER (BP)

Manic-depressive illness, or bipolar mood disorder (BP), is characterized by episodes of elevated mood (mania) and depression and is among the most prevalent and potentially devastating of psychiatric syndromes. The most severe and clinically distinctive forms of BP are BP-I (severe bipolar mood disorder) and SAD-M (schizoaffective disorder manic type), and are characterized by at least one full episode of mania, with or without episodes of major depression (defined by lowered mood, or depression, with associated disturbances in rhythmic behaviors such as sleeping, eating, and sexual activity). A milder form of BP is BP-II, bipolar mood disorder with hypomania and major depression. BP-I often co-segregates in families with more etiologically heterogeneous syndromes, such as unipolar major depressive disorder (MDD), which is a more broadly defined phenotype. See McInnes, L.A. and Freimer, N.B., Mapping genes for psychiatric disorders and behavioral traits, Curr. Opin. in Genet. and Develop., 5:376-381 (1995).

## TREATMENT OF INDIVIDUALS WITH BIPOLAR MOOD DISORDER

An estimated 2-3 million people in the United States are affected by BP-I. Currently, individuals are typically evaluated for bipolar mood disorder using the clinical criteria set forth in the most current version of the American Psychiatric Association's

5 Diagnostic and Statistical Manual of Mental Disorders (DSM). Many drugs have been used to treat individuals diagnosed with bipolar mood disorder, including lithium salts, carbamazepine and valproic acid. However, none of the currently available drugs is able to treat every individual diagnosed with severe BP-I (termed BP-I) and drug treatments are effective in only approximately 60-70% of individuals diagnosed

10 with BP-I. Moreover, it is currently impossible to predict which drug treatments will be effective in particular BP-I affected individuals. Commonly, upon diagnosis affected individuals are prescribed one drug after another until one is found to be effective. Early prescription of an effective drug treatment is critical for several reasons, including the avoidance of extremely dangerous manic episodes and the risk

15 of progressive deterioration if effective treatments are not found. Also, appropriate treatment may prevent depressive episodes in BP-I individuals; these episodes are also dangerous and are characterized by a high suicide rate. The high prevalence of the disorder, together with frequent occurrence of hospitalizations, psychosocial impairment, suicide and substance abuse, has made BP-I a major public health

20 concern.

### Genetic Basis for Bipolar Mood Disorder

Mapping genes for common diseases believed to be caused by multiple genes, such as BP-I, may be complicated by the typically imprecise definition of phenotypes, by

25 etiologic heterogeneity and by uncertainty about the mode of genetic transmission of the disease trait. With psychiatric disorders there is even greater ambiguity in distinguishing between individuals who likely carry an affected genotype from those who are genetically unaffected. For example, one can define an affected phenotype for BP by including one or more of the broad grouping of diagnostic classifications

30 that constitute the mood disorders: BP-I, SAD-M, MDD, and BP-II.

Thus, one of the greatest difficulties facing psychiatric geneticists is uncertainty regarding the validity of phenotype designations, since clinical diagnoses are based

solely on clinical observation and subjective reports. Also, with complex traits such as psychiatric disorders, it is difficult to map the trait-causing genes genetically because: (1) the BP-I phenotype doesn't exhibit classic Mendelian recessive or dominant inheritance patterns attributable to a single genetic locus, (2) there may be incomplete penetrance i.e., individuals who inherit a predisposing allele may not manifest the disease; (3) the phenocopy phenomenon may occur, i.e., individuals who do not inherit a predisposing allele may nevertheless develop the disease due to environmental or random causes; (4) genetic heterogeneity may exist, in which case mutations in any one of several genes may result in identical phenotypes.

10 The existence of one or more major genes associated with BP-I and with a clinically similar diagnostic category, SAD-M (schizoaffective disorder manic subtype), is supported by segregation analyses and twin studies (Bertelson et al., 1977; Freimer and Reus, 1992; Pauls et al., 1992). However, efforts to identify the chromosomal location of BP-I genes have yielded disappointing results in that reports of linkage between BP-I and markers on chromosomes X and 11 could not be independently replicated nor confirmed in the re-analyses of the original pedigrees (Baron et al., 1987; Egeland et al., 1987; Kelsoe et al., 1989; Baron et al., 1993). The possible localization of BP genes on chromosomes 18 (pericentromeric region) and 21q has been suggested, but in both cases the proposed candidate region is not well defined and there is equivocal support for either location (Berrettini et al. (1994) Proc. Natl. Acad. Sci. USA, 91, 5918-5921, Murray, J.C., et al. (1994) Science 265, 2049-2054; Pauls et al., Am. J. Hum. Genet. 57:636-643 (1995); Maier et al., Psych. Res. 59:7-15 (1995); Straub et al., Nature Genet., 8:291-296 (1994)). Recent investigations have led to the isolation of chromosome 18-specific brain transcripts which have been suggested to be positional candidates for bipolar disorder (Yoshikawa et al., Am. J. Med. Gen. 74, 140-149 (1997)).

25 Despite abundant evidence that BP has a major genetic component, linkage studies have not yet succeeded in definitively localizing a BP gene. This is mainly because mapping studies of psychiatric disorders have generally been conducted under a paradigm appropriate for mapping genes for simple Mendelian disorders, namely, using linkage analysis in the expectation of finding high lod scores that definitively signpost the location of disease genes. The follow up to early BP linkage studies,

however, showed that even extremely high lod scores at a single location can be false positives. See Egeland, et al., Nature 325:783-787 (1987); Baron et al., Nature 326:289-292 (1987); Kelsoe et al., Nature, 342:238-243 (1989); and Baron et al., Nature Genet. 3:49-55 (1993). These earlier studies used largely uninformative  
5 markers and did not use stringent criteria for identifying affected individuals.

### LINKAGE DISEQUILIBRIUM ANALYSIS

Linkage disequilibrium (LD) analysis is a powerful tool for mapping disease genes and may be particularly useful for investigating complex traits. LD mapping  
10 is based on the following expectations: for any two members of a population, it is expected that recombination events occurring over several generations will have shuffled their genomes, so that they share little in common with their ancestors. However, if these individuals are affected with a disease inherited from a common ancestor, the gene responsible for the disease and the markers that immediately  
15 surround it will likely be inherited without change, or IBD ("identical by descent"), from that ancestor. The size of the regions that remain shared (i.e. IBD) are inversely proportional to the number of generations separating the affected individuals and their common ancestor. Thus, "old" populations are suitable for fine scale mapping and recently founded ones are appropriate for using LD to roughly localize  
20 disease genes more approximately (Houwen et al., 1994, in particular Fig. 3 and accompanying text). Because isolated populations typically have had a small number of founders, they are particularly suitable for LD approaches, as indicated by several successful LD studies conducted in Finland (de la Chapelle, 1993).

LD analysis has been used in several positional cloning efforts (Kerem et al.,  
25 1989; MacDonald et al., 1992; Petrukhin et al., 1993; Hastbacka et al., 1992 and 1994), but in each case the initial localization had been achieved using conventional linkage methods. Positional cloning is the isolation of a gene solely on the basis of its chromosomal location, without regard to its biochemical function. Lander and Botstein (1986) proposed that LD mapping could be used to screen the human genome  
30 for disease loci, without conventional linkage analyses. This approach was not practical until a set of mapped markers covering the genome became available

(Weissenbach et al., 1992). The feasibility of genome screening using LD mapping is now demonstrated by the applicants.

Identification of the chromosomal location of a gene responsible for causing severe bipolar mood disorder can facilitate diagnosis, treatment and genetic  
5 counseling of individuals in affected families.

Due to the severity of the disorder and the limitations of a purely phenotypic diagnosis of BP-I, there is a tremendous need to subtype individuals with BP-I genetically to confirm clinical diagnoses and to determine appropriate therapies based on their genotypic subtype.

10

### SUMMARY OF THE INVENTION

The present invention comprises using genetic linkage and haplotype analysis to identify an individual having a bipolar mood disorder gene on the short arm of chromosome 18. In addition, the present invention provides markers linked to a gene  
15 responsible for susceptibility to bipolar mood disorder that will enable researchers to focus future analysis on that small chromosomal region and will accelerate the sequencing of a bipolar mood disorder gene located at 18p.

The present invention provides, for the first time, a localization of a BP-I susceptibility locus to a 300 to 500 kb region of the short arm of chromosome 18.

20 The present invention is directed to methods of detecting the presence of a bipolar mood disorder susceptibility locus in an individual, comprising analyzing a sample of DNA for the presence of a DNA polymorphism on the short arm of chromosome 18 between SAVA5 and ga203, wherein the DNA polymorphism is associated with a form of bipolar mood disorder. The invention includes the use of  
25 genetic markers in the roughly 500 kb region between the SAVA5 locus and the ga203 locus, inclusive, to diagnose bipolar mood disorder genetically in individuals and to confirm phenotypic diagnoses of bipolar mood disorder. Preferably, the sample of DNA is analyzed for the presence of a DNA polymorphism on the short arm of chromosome 18 in the roughly 300 kb region between D18S1140 and W3422.

30

In a further embodiment, the invention provides methods of classifying subtypes of bipolar mood disorder by identifying one of more DNA polymorphisms

located within the 500 kb region between SAVA5 and ga203 loci, inclusive, on the short arm of chromosome 18 and analyzing DNA samples from individuals phenotypically diagnosed with bipolar mood disorder for the presence or absence of one or more of said DNA polymorphisms. Preferably, the sample of DNA is  
5 analyzed for the presence or absence of one or more of said DNA polymorphisms in the roughly 300 kb region between D18S1140 and W3422 on the short arm of chromosome 18.

In yet a further embodiment, the methods of the invention include a method of treating an individual diagnosed with bipolar mood disorder comprising identifying  
10 one or more DNA polymorphisms located within the 500 kb region of chromosome 18 between SAVA5 and ga203, analyzing DNA samples from individuals phenotypically diagnosed with bipolar mood disorder for the presence or absence of one or more of the DNA polymorphisms, and selecting a treatment plan that is most effective for individuals having a particular genotype within the 500 kb region of  
15 chromosome 18 between SAVA5 and ga203. Preferably, the sample of DNA is analyzed for the presence or absence of one or more DNA polymorphisms in the roughly 300 kb region between D18S1140 and W3422 on the short arm of chromosome 18.

## 20 BRIEF DESCRIPTION OF THE DRAWINGS

**FIG. 1** is a pedigree chart showing two families, CR001 and CR004. Affected individuals are denoted by black symbols, deceased individuals by a diagonal slash. A schematic of each individual's haplotype (where available) is shown below the ID number. Recombinations are denoted by "-x"; consanguineous marriages by  
25 a double bar, and the conserved haplotype as black shading within the haplotype bars. The larger conserved region for CR004 is stippled, the larger conserved region for CR001 is indicated by a dashed outline. An "I" underneath the haplotype bars indicates inferred haplotype. A "?" indicates phase is uncertain. The connection between CR001 and CR004, dating to an 18th Century founding couple, is indicated  
30 by the dashed lines joining individuals III-6 and I-4.

FIG. 2 is a table of lod scores for markers covering the entire human genome that exceeded the arbitrary coverage thresholds. Lod scores are shown for two markers on chromosome 18: D18S59 and D18S1105.

5        FIG. 3 depicts the extent of marker coverage used in the pedigree genome screening study for each chromosome. Coverage is defined as regions for which a lod score of at least 1.6 would have been detected (in the combined data set) for markers truly linked to BP-I under the model employed. Areas that remain uncovered (at this threshold) are unshaded. Markers for which lod scores were obtained that exceeded  
10       the empirically determined coverage thresholds in CR001, CR004, or the combined data set, are shown at their approximate chromosomal location. The symbols to the right of the chromosome indicate the thresholds exceeded at that marker: a circle signifies that the lod score at a marker exceeded the threshold of 0.8 in CR001, a diamond signifies that the lod score exceeded the threshold of 1.2 in CR004, and a  
15       star signifies that the lod score exceeded the threshold of 1.6 in the combined data set.

      FIGS. 4A and 4B depicts the Lod score for the maximum likelihood estimate of theta in the combined sample for the 473 microsatellite markers typed in the  
20       pedigree genome screen. The MLEs of theta were appointed to the following categories:  $\theta < 0.10$ ;  $0.10 \leq \theta \leq 0.40$ ;  $\theta \geq 0.40$ . Note that the scale for the x-axis (distance from pter) changes with chromosomes.

      FIG. 5 is a portion of an integrated map of the 5 cM 18pter region of  
25       chromosome 18.

      FIGS. 6A, 6B and 6C are a list of markers on chromosome 18, with map positions noted.

30       FIG. 7 describes 18p allele frequencies for disease chromosomes (aff 105) versus nontransmitted chromosomes (ntrans) and samples from a control population of Costa Rican students and their parents (control). The name of each marker used

in this study is indicated on the left. The second column of numbers refers to allele length in base pairs.

FIG. 8 depicts haplotype analysis of individuals affected with BP-I. The column labelled 18p refers to the patient identifier, and each patient identifier is repeated with 2 rows to indicate allele results with each of the patient's two copies of chromosome 18. The columns labelled "PANR" and "MANR" refer to the paternal and maternal identifiers, respectively, associated with the particular patient, other than 0, 1 and 2, which indicate that parental samples were not available. The column headings to the right of "PANR" and "MANR" columns represent names of specific markers in the 18p region that were used in the haplotype analysis. The markers are listed in the order they appear on chromosome 18. The allele length (in base pairs) is indicated under the column heading each marker for a particular patient. In the column to the immediate right of each marker column, a "1" indicates that the phase is known, i.e., that it is known whether a particular allele is inherited from the paternal or maternal chromosome, and a "0" indicates that the phase is not definitely known. The shaded horizontal bars depict haplotypes of at least three markers which include a 154 allele length at D18S59, other than patients 218, 225, 232, 234, 311, 314 and 458, where the shaded region depicts small sections that do not have the 154 allele at D18S59. The lightly shaded regions depict uncertainty as to whether the individual has the affected haplotype, as the phase is not known with certainty. In addition, the presence of an allele length of 232 (or 234) with marker ta201 is thought to result from a highly mutable allele and may not be distinct from the 230 allele. Similarly, the 202 allele at ca212 may not be distinct from the 200 allele at ca212. Patients 246, 247, 248, 311, 316, 367, 384, 501, 531, 587, 536, 684, 667 and 669 exhibit a 242, 244, 250, 252 or 214 allele at marker ta201 which indicates a potential marker location. Patients 488, 435 and 236 exhibit haplotypes that are distinct from the pedigrees that were analyzed.

FIG. 9 depicts haplotype analysis of nontransmitted chromosomes from parents of individuals affected with BP-I. The labels "ERSN" and "KID" refer to the parental and patient identifiers, respectively. As above, allele length is provided in

base pairs below each marker with an indication as to whether phase was known (1) or not known (0) given to the right of these values. The markers, shading and allele characteristics described for Figure 8 also apply to this figure.

5

FIG. 10 depicts haplotype analysis of control samples obtained from an unscreened population of students of the University of Costa Rica and their parents representing the general population. Identifiers are provided in the column headed "cont", allele length and phase determination given in the remainder of the table. The markers, shading and allele characteristics described for Figure 8 also apply to this figure. Complete data for all markers are not given as indicated by blank boxes, or the terms "miss" or "missing".

FIG. 11 depicts Ancestral Haplotype Reconstruction results in disease chromosomes.

15

#### DESCRIPTION OF SPECIFIC EMBODIMENTS

The recent availability of highly polymorphic, genetically mapped markers covering the human genome (Weissenbach, J., et al. (1996) Nature 359, 794-801, Murray, J.C., et al. (1994) Science 265, 2049-2054, Gyapay, G., et al. (1994) Nature Genet 7,246-339) has allowed the development of a multi-stage paradigm for mapping genes for complex traits. In the first stages, complete genome screening (e.g. through lod score analysis) is used to identify possible localizations for disease genes. Subsequently, the regions highlighted by the screening study are more intensively investigated to confirm the initial localizations and delineate clear candidate regions. Finally, fine mapping methods (such as haplotype or linkage disequilibrium (LD) analysis) or candidate gene approaches are used for positional cloning of disease genes.

25

Our genome screening study for BP employed the following strategies. Unlike previous genetic studies of BP, only those individuals with the most severe and clinically distinctive forms of BP (BP-I and schizoaffective disorder manic type, SAD-M) were considered as affected, rather than including those diagnosed with a milder

30

form of BP (BP-II) or with unipolar major depressive disorder (MDD). Two large pedigrees (CR001 and CR004) were selected from a genetically homogeneous population, that of the Central Valley of Costa Rica (as described in Escamilla, M.A., et al., (1996) Neuropsychiat. Genet. 67, 244-253, and in Freimer, N.B., et al. (1996) Neuropsychiat. Genet. 67, 254-263, both incorporated by reference herein). The entire human genome was screened for linkage using mapped microsatellite markers and a model for genetic analysis in which most of the linkage information was derived from affected individuals. The goal of this stringent linkage analysis was to identify all regions potentially harboring major genes for BP-I in the study population. Empirically determined lod score thresholds (using linkage simulation analyses) were derived, to suggest regions worthy of further investigation.

Identification of all suggestive regions and weighing the relative importance of findings required complete screening of the genome. The coverage approach was developed to gauge the progress of this effort. Conventionally, the thoroughness of genome screening is evaluated by excluding genome regions from linkage under given genetic models. This approach, which is highly sensitive to misspecification of genetic models, may be poorly suited for genome screening studies of complex traits; it is tied to the expectation of finding linkage at a single locus and demonstrating absence of linkage at all other locations in the genome. Additionally, exclusion analyses do not differentiate between genome regions where linkage is not excluded because markers are uninformative in the study population from those in which the genotype data are simply ambiguous. In contrast, the coverage approach is designed for studies aimed at genome screening rather than for studies where the goal is to demonstrate a single unequivocal linkage finding, and it provides explicit data regarding the informativeness of markers in the study pedigrees. Its use lessens the possibility that one would prematurely dismiss a given genome region as being unpromising for further study.

Because the exact genetic length of chromosomes is not clearly established, it is impossible to be certain that one has screened the entire genome. Although we report coverage of about 94 % of the genome (under the 90 %) dominant model) at the thresholds described above, this probably represents an underestimate. The remaining coverage gaps in our study occur predominantly at or near telomeres; as the upper

bound estimates for the length of each chromosome were used, it is likely that the actual coverage gaps in these regions are smaller than our conservative assessment.

The presence of consistently positive lod scores over a given region was considered to be of greater significance than isolated peak lod scores. Such clustering suggests true co-segregation of markers and phenotypes (i.e. alleles are shared identically by descent rather than identically by state) and is more readily observed in analyses of a few large pedigrees (as in our study) than in examination of several smaller families. The data presented herein indicates clustering of positive lod scores in the region of the telomere of 18p.

The genome screen was conducted in two stages. The Stage I screen identified areas suggestive of linkage, so that those areas could be saturated with available markers, and so that regions, referred to as 'coverage gaps', could be pinpointed where markers were insufficiently informative in our sample to detect evidence of linkage. The Stage II screen followed up on regions flanking each marker that yielded peak lod scores approximately equal to or greater than the thresholds used for the coverage calculations, which were deemed regions of interest, and filled in coverage gaps. The results of the complete genome screen (Stages I and II) using 473 markers is described below.

In addition, linkage disequilibrium analysis of an independently collected sample of 48 unrelated BP-I patients was initially conducted. These patients were from the same ancestral population as the patients in the CR001 and CR004 pedigrees. The LD analysis was conducted with markers on the short arm of chromosome 18 (18p), in a 5 centimorgan (cM) region ("5 cM 18pter region") extending from the end of the 18p telomere to a distance of 5 cM along the short arm of chromosome 18 (18p). The LD analysis gave evidence of LD in this region, particularly at marker D18S59 and also at D18S476. LD analysis of further BP-I patients from the CRCV with markers in this 5 cM 18pter region was conducted to confirm and fine map a BP-I gene in this region. This approach, using additional BP-I patients from this CRCV population and additional markers identifies the region of maximum LD and can precisely localize a BP-I susceptibility gene.

Fine mapping of 5 cM 18pter region resulted in the identification of two DNA markers (D18S1140 and W3422) defining the boundaries of BP-I as approximately 300 kb, thus allowing a systematic search for the BP-I gene(s).

5 A conservative approach to linkage analysis was used in that almost all of the information for linkage is derived from individuals with a severe, narrowly defined phenotype. While this approach made it very unlikely that lod scores greater than conventional thresholds of statistical significance (e.g.  $\geq 3$ ) would be obtained, it provided confidence in the robustness of the most suggestive findings.

10 Direct cDNA selection can be used to isolate segments of expressed DNA from the 300 kb region between D18S1140 and W3422 (M. Lovett, J. Kere, L.M. Hinton, *Proc. Natl. Acad. Sci. USA* 88 9628-9632 (1991); Y.-S. Jou *et al.*, *Genomics* 24 410-413 (1994)). By using bacterial artificial chromosomes (BAC) (e.g., commercially available from Research Genetics Inc. Huntsville, Alabama), a group of cDNAs can be identified, and hybridization and PCR-amplification experiments can  
15 be used to determine if these cDNA segments are derived from the 300 kb region.

The cDNAs can then be used to determine whether specific sequences are expressed at lower levels (or not at all) in affected individuals compared to non-carrier individuals. Measurement of mRNA levels in lymphoblastoid cell lines can  
20 be used as an initial screen. The cell lines are prepared by drawing blood from individuals, transforming the lymphoblasts with EBV and growing the immortalized cells in culture. Total RNA and DNA are extracted from the cultured human lymphoblastoid cell lines. Northern blot hybridization is used to determine reduced levels of a specific sequence compared to levels from an unaffected, non-carrier  
25 individual as a result of mutations in the BP-I gene on the chromosomes from these affected individuals which results in decreased levels of mature mRNA and play a primary role in BP-I. Thus, alterations in gene sequences in affected individuals can be determined.

The polymerase chain reaction (PCR) is used to amplify the gene and to  
30 determine its sequence from affected individuals. Sequence comparison with unaffected, non-carrier individuals is carried out to identify polymorphisms in the gene sequence that are responsible for BP-I.

The identification of the biochemical defect that causes BP-I provides a basis for treatments for this disease. In addition, knowledge that certain mutations in the gene are responsible for the disease allows mutation detection tests to be used as a definitive diagnosis for BP-I.

5           Thus, the present invention allows the isolation of a nucleic acid molecule that can be used in the identification of the presence (or absence) of a mutation in the BP-I gene a human and thus can be used in the diagnosis of BP-I or in the genetic counseling of individuals, for example those with a family history of BP-I (although the general population can be screened as well). In particular, it should be noted that  
10           any mutation in the BP-I gene away from the normal gene sequence is an indication of a potential genetic flaw; even so-called "silent" mutations that do not encode a different amino acid at the location of the mutation are potential disease mutations, since such mutations can introduce into (or remove from) the gene an untranslated genetic signal that interferes with the transcription or translation of the gene. Thus,  
15           advice can be given to a patient concerning the potential for transmission of BP-I if any mutation is present. While an offspring with the mutation in question may or may not have symptoms of BP-I, patient care and monitoring can be selected that will be appropriate for the potential presence of the disease; such additional care and/or monitoring can be eliminated (along with the concurrent costs) if there are no  
20           differences from the normal gene sequence. As additional information (if any) becomes available (e.g., that a given silent mutation or conservative replacement mutation does or does not result in BP-I), the advice given for a particular mutation may change. However, the change in advice given does not alter the initial determination of the presence or absence of mutations in the gene causing BP-I.

25           Generally, mutations are identified in the human gene for use in a method of detecting the presence of a genetic defect that causes or may cause BP-I, or that can or may transmit BP-I to an offspring of the human. Initially, the practitioner will be looking simply for differences from the sequence identified as being normal and not associated with disease, since any deviation from this sequence has the potential of  
30           causing disease, which is a sufficient basis for initial diagnosis, particularly if the different (but still unconfirmed) gene is found in a person with a family history of BP-I. As specific mutations are identified as being positively correlated with BP-I (or

its absence), practitioners will in some cases focus on identifying one or more specific mutations of the gene that changes the sequence of a protein product of the gene or that results in the gene not being transcribed or translated. However, simple identification of the presence or absence of any mutation in the gene of a patient will  
5 continue to be a viable part of genetic analysis for diagnosis, therapy and counseling.

The actual technique used to identify the gene or gene mutant is not itself part of the practice of the invention. Any of the many techniques to identify gene mutations, whether now known or later developed, can be used, such as direct  
10 sequencing of the gene from affected individuals, hybridization with specific probes, which includes the technique known as allele-specific oligonucleotide hybridization, either without amplification or after amplification of the region being detected, such as by PCR. Other analysis techniques include single-strand conformation polymorphism (SSCP), restriction fragment length polymorphism (RFLP), enzymatic  
15 mismatch cleavage techniques and transcription/translation analysis. All of these techniques are described in a number of patents and other publications; see, for example, "Laboratory Protocols for Mutation Detection" (1996) Oxford University Press, Editor: Ulf Landegrun.

Depending on the patient being tested, different identification techniques can  
20 be selected to achieve particularly advantageous results. For example, for a group of patients known to be associated with particular mutations of the gene, oligonucleotide ligation assays, "mini-sequencing" or allele-specific oligonucleotide (ASO) hybridization can be used. For screening of individuals who are not known to be associated with a particular mutation, single-strand conformation polymorphism,  
25 total sequencing of genetic and/or cDNA and comparison with standard sequences are preferred.

In many identification techniques, some amplification of the host genomic DNA (or of messenger RNA) will take place to provide for greater sensitivity of analysis. In such cases it is not necessary to amplify the entire gene, merely the part  
30 of the gene or the specific location within the gene that is being detected. Thus, the method of the invention generally comprises amplification (such as via PCR) of at

least a segment of the gene, with the segment being selected for the particular analysis being conducted by the diagnostician.

The patient on whom diagnosis is being carried out can be an adult, as is usually the case for genetic counseling, or a newborn, or prenatal diagnosis can be carried out on a fetus. Blood samples are usually used for genetic analysis of adults  
5 or newborns (e.g., screening of dried blood on filter paper), while samples for prenatal diagnosis are usually obtained by amniocentesis or chorionic villus biopsy.

Prior to the present invention, affected individuals were prescribed one drug after another until one was found to be effective. As BP-I was diagnosed using  
10 clinical criteria, no correlation between using a particular drug and its efficacy in a given case was observed. As a result of the present invention, BP-I subtypes can be diagnosed at the molecular level and effective treatment predicted.

For example, lithium salts, carbamazepine and valproic acid have all been prescribed for BP-I affected individuals with serendipitous results. An individual can  
15 now be diagnosed with bipolar mood disorder by analyzing genetic material from that individual for the presence or absence of one or more nucleic acid mutations as described above. As a result of this diagnosis at the molecular level, an effective treatment can be determined by collecting data to obtain a statistically significant correlation of a particular treatment with the different subtypes of BP-I. Thus, the  
20 practitioner is able to select a specific drug for the treatment of a particular sub-type of BP-I and does not merely rely on trial and error.

Alternatively, the full-length normal genes for BP-I from humans, as well as shorter genes that produce functional proteins, can be used to correct BP-I in a human patient by supplying to the human an effective amount of a gene product of the human  
25 gene, either by gene therapy or by *in vitro* production of the protein followed by administration of the protein. It should be recognized that the various techniques for administering genetic materials or gene products are well known and are not themselves part of the invention. The invention merely involves supplying the genetic materials or proteins identified as a result of the present invention in place of the  
30 genetic materials or proteins previously administered. For example, techniques for transforming cells to produce gene products are described in U.S. Patent No. 5,283,185 entitled "Method for Delivering Nucleic Acid into Cells," as well as in

numerous scientific articles, such as Felgner et al., "Lipofection: A Highly Efficient, Lipid-Mediated DNA-Transfection Procedure," *Proc. Natl. Acad. Sci. U.S.A.*, 84, 7413-7417 (1987); techniques for in vivo protein production are described in, for example, Mueller et al., "Laboratory Methods - Efficient Transfection and Expression of Heterologous Genes in PC12 Cells," *DNA and Cell Biol.*, 9(3), 221-229 (1990).

Administration of proteins and other molecules to overcome a deficiency disease is well known (e.g., administration of insulin to correct for high blood sugar in diabetes) that further discussion of this technique is not necessary. Some modification of existing techniques may be required for particular applications, but those modifications are within the skill level of the ordinary practitioner using existing knowledge and the guidance provided in this specification.

The invention now being generally described, the following examples are provided for purposes of illustration only and are not to be considered to limit the invention.

## EXAMPLES

### PEDIGREES

Two independently ascertained Costa Rican pedigrees (CR001 and CR004) were chosen because they contained a high density of individuals with BP-I and because their ancestry could be traced to the founding population of the Central Valley of Costa Rica. The current population of the Central Valley (consisting of about two million people) is predominantly descended from a small number of Spanish and Amerindian founders in the 16th and 17th centuries (Escamilla, M.A., et al., (1996) *Neuropsychiat. Genet.* 67, 244-253). Studies of several inherited diseases have confirmed the genetic isolation of this population (Leon, P., et al. (1992) *Proc. Natl. Acad. Sci. USA.* 89, 5181-5184; Uhrhammer, N., et al. (1992) *Am. J. Hum. Genet.* 57, 103-111). An extensive description of pedigrees CR001 and CR004 has been published (Freimer, N.B., et al. (1996) *Neuropsychiat. Genet.* 67, 254-263). In the course of the study, two links between these pedigrees were discovered. However, the families were analyzed separately because these links were

discovered after the simulation analyses were completed and after the genome screening study had been initiated.

5 All available adult members of these families were interviewed in Spanish using the Schedule for Affective Disorders and Schizophrenia Lifetime version (SADS-L) (Endicott, J. et al, (1978) Arch. Gen. Psych. 35, 837-844). Individuals who received a psychiatric diagnosis were interviewed again in Spanish by a research psychiatrist using the Diagnostic Interview for Genetic Studies (DIGS) (Nurnberger, J.L. et al. (1994) Arch. Gen. Psychiat. 51, 849-859). This recently developed instrument is similar to, but more detailed than SADS-L. The interviews and medical  
10 records were then reviewed by two blinded best estimators who reached a consensus diagnosis. The diagnostic procedures are described in detail in Freimer, N.B., et al. (1996) Neuropsychiat. Genet. 67, 254-263 (incorporated by reference herein).

#### UNRELATED CRCV BP-I PATIENT STUDY

15 BP localizations obtained through the CRCV pedigree studies were confirmed by genotyping an independently collected sample of 48 unrelated BP-I patients from the CRCV. In this fine mapping LD analysis, 48 unrelated BP-I patients from the CRCV were identified and genotyped using microsatellite markers spaced at narrow intervals across chromosome 18. As these patients are descended from the same  
20 ancestral population as the patients in the pedigrees previously studied (CR001 and CR004), many of them should share disease susceptibility alleles inherited identically by descent (IBD) from one or a few common ancestors, and linkage disequilibrium (LD) should be present at marker loci surrounding the disease genes.

The sample of 48 BP-I patients included 25 women and 23 men who were  
25 recruited from psychiatric hospitals and clinics in the CRCV. These patients were ascertained only on the basis of diagnosis and CV ancestry, and were not selected on the basis of history of BP illness in family members. A structured interview of each patient was conducted by a psychiatrist, and medical and hospital records were collected. Ascertainment and diagnostic procedures were as described above.  
30 However, in order to lessen further the probability of phenocopies among this unrelated sample, for which we lacked pedigree information, the affected phenotype was defined even more narrowly than in the pedigree study. Individuals considered

affected in this study had to have suffered at least two disabling episodes of mania (requiring hospitalization) and a first onset of the illness before age 45.

Genealogical research on each of the 48 BP-I patients confirmed that on average, 70% of their great-grandparents were born in the CRCV. Individuals whose great-grandparents were born in the CRCV were considered likely to be descended from the original Spanish and Amerindian founders of the CRCV. Genealogical research showed that 2 patients are first cousins and the remaining 46 have no relationship within the past 4 generations.

#### 10 GENOTYPING PEDIGREE STUDIES

Linkage simulations were used to select the most informative individuals from pedigrees CR001 and CR004 for genotyping studies (Freimer, N.B., et al. (1996) Neuropsychiat. Genet. 67, 254-263). Under a 90% dominant model, simulation analyses with these individuals suggested that evidence of linkage would likely be detected (e.g. a probability of 92% of obtaining lod > 1.0 in the combined data set) using markers with an average heterozygosity of 0.75 spaced at 10 cM intervals (as discussed in Freimer, N.B., et al. (1996) Neuropsychiat. Genet. 67, 254-263). For the Stage I screen, the most polymorphic markers (307 in total) were chosen, placed at approximately 10 cM intervals on the 1992 Genethon map (Houwen, R., et al. (1992) Nature 359, 794-801). These markers were then supplemented by a small number of markers from the Cooperative Human Linkage Center (CHLC) public database. For the Stage II screen, 166 markers were added from newer Genethon and CHLC maps as they became available (Murray, J.C. et al. (1994) Science 265, 2049-2054, Gyapay, G., et al. (1994) Nature Genet. 7,246-339) and from the public database of the Utah Center for Genome Research, all of which are publicly available. DNA samples (from individuals in the CEPH families) that were used for size standards for Genethon and CHLC markers were included in the experiments to permit comparison of allele sizes between members of the CRCV population and individuals in the CEPH database. Genotyping procedures were as described previously (DiRienzo, A. et al. (1994) Proc. Natl. Acad. Sci. USA 91, 3166-3170 (incorporated by reference herein)). Briefly, one of the two PCR primers was labeled radioactively using a polynucleotide kinase and PCR products were run on

polyacrylamide gels. Autoradiographs were scored independently by two raters. Data for each marker were entered into the computer database twice and the resultant files were compared for discrepancies.

## 5 GENOTYPING OF UNRELATED BP-I CRCV PATIENTS

Twenty-seven markers were used to genotype all 48 individuals (as well as 53 individuals used to establish genetic phase) at approximately 5 cM intervals along the entire chromosome 18. It was hypothesized that such a screen would permit the evaluation of evidence in the 18pter region and also to investigate other regions on  
10 chromosome 18 in which linkage to BP has been suggested by other groups in other populations. For each individual, two-marker haplotypes in each of the 26 inter-marker intervals were investigated. For 38 of the 48 BP-I patients, genotypes of parents or children were available to assist in phase determination. Because of phase ambiguities in the remaining 10 individuals, minimal and maximal two-marker  
15 haplotype sharing was evaluated as follows: (1) Minimal: the number of individuals (and chromosomes) who definitely shared a chromosomal segment defined by a particular pair of alleles (phase known chromosomes) and (2) Maximal: the number of individuals (and chromosomes) who could possibly share a chromosomal segment defined by a particular pair of alleles (includes phase unknown chromosomes). The  
20 threshold used to identify areas of high IBD sharing of chromosomes in this initial screen was designated as maximal sharing of a two-marker haplotype by 50 % or more of the 48 individuals (or 25 % or more of the 96 chromosomes).

Arbitrary thresholds were designated to identify possible areas of high IBD sharing among the 48 patients. Eight of the 26 regions passed this screen. Within  
25 each of these 3 regions, one to three additional markers were typed to permit detection of LD, if present, over regions of one to two cM.

A total of 42 chromosome 18 markers were used to genotype the study sample:

30 D18S1140, D18S59, D18S476, D18S481, D18S391, D18S452, D18S843, D18S464,  
D18S1153, D18S378, D18S53, D18S453, D18S40, D18S66, D18S56, D18S57,  
D18S467, D18S460, D18S450, D18S474, D18S69, D18S64, D18S1134, D18S1147,  
D18S60, D18S68, D18S55, D18S477, D18S61, D18S488, D18S485, D18S541,

D18S870, D18S469, D18S874, D18S380, D18S1121, D18S1009, D18S844, D18S554, D18S461, D18S70 (from pter to qter). Of these 42 markers, four are located within the 5 cM 18pter region extending from the telomere of 18p to marker D18S481 (inclusive), which is approximately 5 cM from the telomere of 18p. This region is referred to as the 5 cM 18pter region. The four markers tested in the 5 cM 18pter region are: D18S59, D18S1140, D18S476 and D18S481.

For each marker the likelihood that a particular allele (or alleles) is over-represented on disease chromosomes, as compared to non-disease chromosomes was evaluated. The results of this likelihood test provide a conservative but powerful measure of LD between two loci.

#### PEDIGREE STATISTICAL ANALYSES

Two-point linkage analyses were performed for all markers. Marker allele frequencies were estimated from the combined data set with correction for dependency due to family relationships (Boehnke, M. (1991) Am. J. Hum. Genet. 48, 22-25). The linkage analyses for Stages I and II included the 65 individuals who were genotyped as well as an additional 65 individuals who had been diagnostically evaluated but not genotyped. Only individuals with BP-I were considered affected with the exception of two persons, one in each family, who carry diagnoses of schizoaffective disorder manic type (SAD-M). The SAD-M individuals were included as affected because BP-I and SAD-M are often difficult to distinguish from each other based on their clinical presentation and course of illness (Goodwin, F.K. et al. (1990) in Manic Depressive Illness (Oxford University Press, New York), pp. 373-401; Freimer, N.B et al. (1993) in The Molecular and Genetic Basis of Neurological Disease, pp. 951-965; Freimer, N.B. et al. (1996) Neuropsychiat. Genet. 67, 254-263; and Freimer, N.B. et al (1996) Nature Genetics 12:436-441, all incorporated by reference herein). In all, 20 individuals were designated as affected within CR004 (Copeman, J.B., et al. (1995) Nature Genet. 9, 80-85 available for genotyping) and 10 individuals from CR001 (Kelsoe, J.R. et al. (1989) Nature 342, 238-243 available for genotyping). The phenotype for all other individuals was designated as unknown except for 17 individuals who were designated as unaffected because they had been thoroughly clinically evaluated, showed no evidence of any psychiatric disorder, and

were well beyond the age of risk (50) for BP-I (linkage simulation studies indicated that these unaffected individuals contributed little information to the linkage analysis).

Linkage analyses were performed using a nearly dominant model (assuming penetrance of 0.81 for heterozygous individuals of 0.9 for homozygotes with the disease mutation). This model was chosen from five different single-locus models (ranging from recessive to nearly dominant) due to its consistency with the segregation patterns of BP in the two pedigrees and because it had demonstrated the greatest power to detect linkage in simulation studies (Freimer, N.B., et al. (1996) *Neuropsychiat. Genet.* 67, 254-263). Based on Costa Rican epidemiological surveys Escamilla, M.A., et al., (1996) *Neuropsychiat. Genet.* 67, 244-253, the population prevalence of BP-I was assumed to be 0.015 (and thus the frequency of the disease allele was assumed to be 0.003)(based on epidemiological surveys in Costa Rica; Adis, G. (1992) "Disordenes mentales en Costa Rica: Observaciones Epidemiologicas," (San Jose, Costa Rica: Editorial Nacional de Salud y Seguridad Social)). The frequency of BP-I in individuals without the disease allele was conservatively set at 0.01 which effectively specified a population phenocopy rate of 0.67 (i.e., an affected individual in the general population has a 2/3 probability of being a phenocopy). For multiply affected families, the probability that a gene segregates is highly increased, which implies that affected individuals in our study pedigree have a lower probability to be phenocopies than affected individuals in the general population, particularly those with several affected close relatives (the exact probabilities are dependent on the degree of relationship between patients and the number of intervening unaffected individuals). These parameters were chosen to ensure that most of the linkage information derives from affected individuals. The rationale for selecting these parameters and results of analyses that demonstrate the conservatism of this model are described by Freimer, N.B., et al. (1996) *Neuropsychiat. Genet.* 67, 254-263. The LINKAGE package (Lathrop et al., (1984) *Proc. Natl. Acad. Sci. USA* 81, 3443-3446) was used for lod score analysis and to obtain maximum likelihood estimates of the marker allele frequencies, taking into account the existing family relationships (see Boehnke, *Am. J. Hum. Gent.* 48, 22-25 (1991)).

### UNRELATED BP-I CRCV PATIENT STATISTICAL ANALYSES

A likelihood test of disequilibrium (J. Terwilliger, Am. J. Hum. Genet. 56, 777 (1995)) was used to estimate a single parameter,  $\lambda$ , that quantifies the over-representation of marker alleles on disease chromosomes as compared to non-disease chromosomes. We chose this method of analysis over another commonly used disequilibrium analysis method, the transmission disequilibrium test (TDT, R. Spielman et al., Am. J. Hum. Genet. 52, 506 (1993)) because data from all 48 BP-I patients could be used in the likelihood approach. Effective use of the TDT requires phase-known, heterozygous parental chromosomes. We do not have parental genotypes for 20 of the 48 BP-I patients. Simulations indicated that with our data, the likelihood test of disequilibrium would be more powerful than the TDT.  $\lambda$  has been shown to be a superior measure for LD fine mapping, compared to other frequently used measures, because it is directly related to the recombination fraction between the disease and the marker loci. Non-disease chromosomes were chosen from the phase-known chromosomes of parents, spouses and children of affected individuals, if available. Designation of chromosomes of family members as non-disease in a disorder such as BP-I, which is not fully penetrant, necessitates specifying a model of disease transmission. The same model of transmission was employed in this LD likelihood test as was used in the initial genome screen of the pedigrees CR001 and CR002 described herein. One parameter was specified differently from the genome screen: the phenocopy rate was set to zero in the LD likelihood analysis. A phenocopy rate was not specified in the transmission model because the effect of phenocopies will be "absorbed" by the  $\lambda$  parameter, in that presence of phenocopies in our sample will serve to erode the association between marker alleles and disease, and hence reduce the estimate of  $\lambda$ .

### COVERAGE

To access coverage for a marker, the number of informative meioses at the estimated recombination fraction was calculated using the estimate of the variance (the inverse of the information matrix) (Petrukhin, K.E. et al. (1993) Genomics 15, 76-85). Alternatively, when the estimated frequency of recombination was close to 0 or 1, Edwards' equation was applied to calculate the equivalent number of observations

(Edwards, J.H. (1971) *Ann. Hum. Genet.* 34, 229-250). These meioses represent the amount of linkage information provided by the marker, given the pedigree structure and the genetic model applied. Linkage to the marker in question was then assumed and the lod score that would be observed as a disease gene is hypothetically moved in increments away from that marker was calculated. All regions around a marker that would have generated a lod score that exceeded our thresholds for possible linkage (0.8 in CR001, 1.2 in CR004, and 1.6 in the combined data) were considered covered. These lod score thresholds were derived from simulation analyses showing the expected distribution of lod scores under linkage and non-linkage (Freimer, N.B., et al. (1996) *Neuropsychiat. Genet.* 67, 254-263, and approximately represent a result that is 250 times more likely to occur in linked simulations than in unlinked simulations. Coverage maps were constructed (FIG. 1) by superimposing the regions covered by each marker on the genetic map of each chromosome. At the end of the Stage II screen, a total of 473 microsatellite markers had been typed with genome coverage (in the combined data set) of over 94%. Possible coverage gaps are indicated by unshaded areas and are mainly concentrated near telomeres. Because the coverage calculations make use of marker informativeness within the pedigrees, the coverage approach thus permits detection of instances where markers with expected high heterozygosities are uninformative in our data set.

#### PEDIGREE LINKAGE ANALYSIS RESULTS

Of the 473 microsatellites analyzed with two-point linkage tests, 23 markers exceeded the empirically determined thresholds designated for the coverage calculations (in either CR001, CR004, or in the combined data set). The location of these markers, the peak lod scores obtained in each family and in the combined data set, and the maximum likelihood estimate of the recombination fraction (0) at which these lod scores were observed are indicated in Table 1. The approximate chromosomal locations of these markers are also depicted in FIG. 1. The distribution of lod scores (for the maximum likelihood estimate of 0 in the combined data set) across the genome is displayed by chromosome in FIG. 2.

The threshold was exceeded for pedigree CR001 in two adjacent markers near the 18p telomere (D18S59 and D18S1105), but CR004 displayed no suggestion of linkage in this region.

In the genome screen, the highest lod score observed for family CR001 alone was at D18S59 (1.32 at  $\theta=0.0$ ), located near pter. All affected members of CR001 shared alleles at markers in the 18pter region.

#### UNRELATED BP-I CRCV PATIENT STUDY RESULTS

Out of the forty-two markers tested, eight displayed evidence of over-representation of a particular allele on disease chromosomes. Eight of the 42 markers had  $-2*\ln(\text{likelihood ratio})$  statistics  $>1.0$ . Three other markers had  $-2*\ln(\text{likelihood ratio})$  statistics  $>0$  and  $<0.62$ . The results are shown in Table I:

Table I

Marker	Allele Size	Frequency on non-disease Chromosomes	Frequency on Disease Chromosomes
D18S59	154	0.121	0.572
D18S476	271	0.470	0.771
D18S467	172	0.384	0.693
D18S61	177	0.074	0.326
D18S485	182	0.237	0.586
D18S870	179	0.405	0.657
D18S469	234	0.128	0.450
D18S1121	168	0.171	0.553

Evidence for association was found at markers located near the telomere of the short arm of chromosome 18. D18S59 displayed the strongest evidence for LD ( $-2*\ln(\text{likelihood ratio})$  of 8.3,  $p=0.002$ ) of all the chromosome 18 markers tested. An adjacent marker, D18S476 ( $-2*\ln(\text{likelihood ratio})$  of 1.3), also provided evidence of LD. In our genome screening pedigree study we observed the single highest lod score for pedigree CR001 of any marker in the entire

genome at D18S59. Furthermore, the alleles at D18S59 and D18S476 that are over-represented among the BP-I patients from the population sample (154 b.p. and 271 b.p. respectively) are observed in all BP-I patients from pedigree CR001.

5        The LD and pedigree findings in the 5 cM 18pter region denote a clearly delineated region that contains a BP-I susceptibility locus. This region is distinct from other regions on chromosome 18 that have been suggested as linked to mood disorder phenotypes (more broadly defined than BP-I). See FIG. 6A, 6B, 6C. In contrast to previous reports by Berrettini et al. and Stine et al., suggesting possible  
10       linkage between mood disorder and markers in the pericentromeric region of chromosome 18, our results did not show any evidence for association of BP-I with any pericentromeric markers (D18S378, D18S53, D18S453 or D18S40).

#### IDENTIFICATION OF NEW MARKERS FROM THE 5 CM 18PTER REGION

15        Cloned human genomic DNA covering the target region is assembled. Microsatellite sequences from these clones are identified. A sufficient area around the repeat to enable development of a PCR assay for genomic DNA is sequenced, and it is confirmed that the microsatellite sequence is polymorphic, as several uninformative microsatellites are expected in any set. Several methods have been  
20       routinely used to identify microsatellites from cloned DNA, and at this time no single one is clearly preferable (Weber, 1990, Hudson et al., 1992). Most of these require screening an excessive number of small insert clones or performing extensive subcloning using clones with larger inserts.

      New strategies have recently been developed which permit the use of the  
25       several different microsatellites to be found within a single large insert clone without requiring extensive subcloning. A method for direct identification of microsatellites from yeast artificial chromosomes (YACs) provides several new markers from the target region. This procedure is based on a subtractive hybridization step that permits separation of the target DNA from the vector  
30       background. This step is useful because the human DNA (the YAC) constitutes only a small proportion of the total yeast genomic DNA.

YAC clones (with inserts averaging about 750 Kb of human genomic DNA) that span the 5 cM 18pter region have already been identified by the CEPH/Généthon consortium (Cohen et al., 1993) and are publicly available. The markers from YACs that have been mapped to portions of the candidate region that are not well represented by currently available markers are first isolated. By typing these markers in the families and the "LD" sample, as described above, it is possible to narrow the candidate region, perhaps to a size of less than one to two cM, thus permitting limitation of the segment in which more extensive mapping efforts are applied.

10 Briefly, the microsatellite identification procedure is performed as follows: A subtractive hybridization is performed using genomic DNA from a target YAC together with an equivalent amount of a control DNA. This procedure separates the YAC DNA from that of the yeast vector. Following the subtraction procedure the subtracted YAC DNA is purified, digested with restriction enzymes and cloned into a plasmid vector (Ostrander et al., 1992). The cloned products of each YAC are screened using a CA(15) oligonucleotide probe (i.e. an oligonucleotide having 15 CA repeats). Each positive clone (i.e. those that contain TG-repeats) is sequenced to identify primers for PCR to genotype the BP-I samples.

20 An alternative approach, based on using a set of degenerate sequencing primers that anneal directly to the repeat sequence, permitting direct thermal cycle sequencing (Browne & Litt, 1992), can also be used.

Once the candidate region is narrowed to a size of less than about 500 to 1000 Kb, a contiguous array (contig) of clones with smaller inserts than YACs, mainly P1 clones, is developed. P1 clones are phage clones specially designed to accommodate inserts of up to 100 Kb (Shepherd et al., 1994).

#### DEVELOPMENT OF A PHYSICAL MAP OF THE 5 cM 18PTER REGION

30 In parallel with the genetic mapping, a physical map of the 5 cM 18pter region is developed. The backbone of this effort is the assembly of contigs of large insert clones. Low resolution contigs for most of the human genome are already available using the YACs developed by CEPH (Cohen et al., 1993). Although these have been individually verified and checked for overlap with other

YACs, there is a high rate of chimerism in the YACs and insufficient evidence to definitively confirm the order of the YACs. In addition, because of their large size these YACs are particularly cumbersome to work with. Nevertheless, they provide a useful framework to start constructing high resolution contigs.

5        Once a candidate region of less than about five cM is delineated, the studies to develop a physical map are commenced. Because of the disadvantages of relying solely on YACs, and because positional cloning is facilitated by the availability of a higher resolution map, contigs are generated using P1 clones once the candidate region is narrowed to less than one Mb, by LD mapping in the  
10        expanded population sample using the new markers identified from the YACs.

      Once a region of 500-1000 Kb or less is defined, physical mapping and cloning are computed using P1 clones rather than YACs, and P1 contigs over such a region are constructed. The P1s are used to identify additional markers for the further positional cloning steps as well as the screening for rearrangements.

15        The starting point of contig construction is the microsatellite sequences and non-polymorphic STSs that derive from the few YACs that surround the genetically determined candidate region. These STSs are used to screen the P1 library. The ends of the P1s are cloned using inverse PCR and used to order the P1s relative to each other. Amplification in a new P1 will indicate that it overlaps  
20        with the previous one. Fluorescent in situ hybridization (FISH) permits ordering of the majority of the P1s (Pinkel, 1988; Lichter, 1991). The original set of P1s serves as building blocks of the complete contig; each end clone is used to re-screen the library and in this way P1s are added to the map.

      From each P1 additional microsatellites are identified as previously  
25        described. This allows further reduction of the candidate region. When the region is narrowed to less than one Mb in size, positional cloning efforts are initiated.

#### FINE MAPPING OF 5CM 18PTER REGION

      In order to delineate further regions of BP-I susceptibility within the 5 cM  
30        18pter region, additional unrelated BP-I patients from the CRCV as well as other populations can be diagnosed and genotyped both with the markers described herein as well as additional markers in the 5 cM 18pter region that are known as

well those yet to be identified. Additional markers are available from the Cooperative Human Linkage Center (CHLC) public database, from newer Genethon and CHLC maps as they become available (Murray, J.C. et al. (1994) Science 265, 2049-2054, Gyapay, G., et al. (1994) Nature Genet. 7,246-339) and  
5 from the public database of the Utah Center for Genome Research (all of which are incorporated by reference herein). The web addresses for Genethon and CHLC are: Genethon ([http://www.genethon.fr/genethon\\_en.html](http://www.genethon.fr/genethon_en.html)), CHLC (<http://gopher.chlc.org/HomePage.html>). These databases are all linked, and one of ordinary skill in the art can readily access the information available from these  
10 databases.

The markers shown in FIG. 6A, from number 1 to 22 or 23 can be used to genotype the CRCV pedigrees and unrelated BP-I patients described herein as well as other BP-I affected individuals and pedigrees. See FIG. 6A (portion of a chromosome 18 map available from the Whitehead Institute, web address:  
15 <http://133.30.8.1:8080/=@@=:www-genome.wi.mit.edu>. (incorporated herein by reference)). The fine mapping techniques described herein in conjunction with the teachings regarding the 5 cM 18pter region can be used to narrow the BP-I susceptibility region further.

The following markers (listed in order of occurrence from the telomere  
20 towards the centromere) were used to delineate regions of BP-I susceptibility within the 5 cM 18pter region: SAVA5, ca211, ca212, D18S1140, D18S59, ca231, ta201, AT201, ca225, w3442, ca213, ga201, ga203, ca219, D18S1105, ca209, ca202, D18S1146, GATA (referred to in the figures as 166d05) and D18S476. The markers SAVA5, D18S1140, D18S59, ta201, at201, w3442,  
25 ga201, ga203, D18S1105, D18S1146, GATA and D18S476 were used in both the haplotype analysis (Figure 8) and the AHR analysis (Figure 11) to delineate the BP-I susceptibility locus to the 500 kb region defined by the markers SAVA5 and ga203 and the 300 kb region defined by D18S1140 and W3422. The other markers were used in both haplotype and the AHR analyses as confirmatory  
30 evidence for the localizations. Blood samples from 105 affected individuals were tested for the presence of marker haplotypes and compared to marker haplotypes detected on the non-transmitted chromosome in samples obtained from the

parent(s) of the affected individuals when available (71 cases) or to markers detected in samples obtained from a control population of students attending the University of Costa Rica (52 samples). The non-transmitted chromosomes are well matched as controls allowing the affected haplotype of the transmitted chromosome to be more easily discerned than through comparison with data obtained from the general population that may contain individuals who carry the affected haplotype but do not exhibit clinical symptoms of bipolar mood disorder.

Figure 7 provides 18p allele frequencies for disease (aff 105) versus nontransmitted (ntrans) chromosomes and samples from the control population of students (control). The name of each marker used in this study is indicated on the left. The second column of numbers refers to allele length in basepairs. This data provides evidence of over-representation of a particular allele on disease chromosomes.

Figure 8 summarizes the results obtained with affected individuals. The column labelled 18p refers to the patient identifier, and each patient identifier is repeated to indicate results with both copies of chromosome 18. The labels "PANR" and "MANR" refer to the paternal and maternal identifier, respectively, associated with the particular patient, other than 0, 1 and 2, which indicate that parental samples were not available. The allele length (base pairs) is indicated under each marker for a particular patient; the length of the horizontal bar in the figure reflects whether haplotypes are IBD or IBS, with IBD haplotypes with common ancestors having longer bars than randomly inherited IBS haplotypes. To the right of each marker, a "1" indicates that the phase is known, i.e., that it is known whether a particular allele is inherited from the paternal or maternal chromosome, and a "0" indicates that the phase is not known for sure. The determination of phase allows the practitioner to conclude that marker alleles are linked in a haplotype on the same disease causing chromosome.

Figure 9 provides similar data for non-transmitted chromosomes obtained from parental samples. Some individuals exhibited the affected haplotype indicating that the parent was homozygous; however, these regions of identity

were typically much shorter than those regions observed in affected individuals, indicating that they were IBS.

Figure 10 similarly provides data for an unscreened population of students from the University of Costa Rica and their parents (52 samples in total).

5 The data demonstrate that there is a lower incidence of the affected haplotype in the general population as compared with Figure 8 and that the affected haplotype is typically shorter similar to the results obtained with non-transmitted chromosomes. However, the results for the general population is less distinctive than that observed for non-transmitted chromosomes in allowing one to map the  
10 affected haplotype.

Comparison of the affected haplotype with non-transmitted chromosome markers indicate that the region of maximal sharing between affected individuals occurs between 1140t and w3442 on chromosome 18. This region encompasses approximately 300 kb.

15 The data was analyzed further using Ancestral Haplotype Reconstruction (AHR), a likelihood method for measuring LD. Data from affected individuals are examined in 2-marker segments. Within each segment, the multinomial likelihood of each of the possible ancestral haplotypes giving rise to the observed sample of disease haplotypes is calculated. This likelihood is calculated assuming  
20 some fraction,  $\alpha$ , of disease chromosomes are associated with this 2-marker segment, and  $(1-\alpha)$  are linked to this segment. These haplotype likelihoods are weighted by the probability of observing that haplotype in the population, and summed to create an overall likelihood for the 2-marker segment. This segment likelihood is compared to the null likelihood, which assumes the disease and  
25 markers are unlinked (and therefore  $\alpha=0$ ), and a LOD score is generated. The LOD score is maximized over the parameter  $\alpha$ . Details of these calculations are presented in Appendix A. The results of this analysis are shown in Figure 11. The percentages given above the diagonal line demarcated by the filled boxes indicate the percentage of disease chromosomes hypothesized to be true  
30 chromosomes from a common founder. For example, 17% of chromosomes obtained from affected individuals have the 18S59 to W3442 region; i.e., as each individual has two chromosome copies, 34% of individuals have this region. The

number above each percentage indicates the LOD score. The numbers given below the diagonal line demarcated by the filled boxes indicate the alleles inherited from a common founder, with the number prior to the dash indicating the allele of the marker identified in the column heading and the number following the dash indicating the allele of the marker identified in the row heading. The marker alleles are referred to as follows:

	MARKER	#	ALLELE LENGTH
	SAVA5	2	229
10	CA211	3	195
	18S1140	2	268
	18S59	4	154
	18S59	6	158
	TA201	2	220
15	TA201	3	230
	CA231	2	186
	CA231	4	202
	AT201	1	170
	AT201	2	178
20	CA225	1	160
	CA225	3	172
	W3442	1	10

Blank boxes indicate no positive evidence for linking the indicated region to the affected chromosome.

#### USE OF P1 CLONES TO IDENTIFY CANDIDATE cDNAs FOR SCREENING FOR MUTATIONS IN THE DNA OF BP-I PATIENTS

The P1 clones described above are used to identify candidate cDNAs. The candidate cDNAs are subsequently screened for mutations in DNA from BP-I patients. From the minimal candidate region defined by genetic mapping experiments a segment is left that is sufficiently large to contain multiple different genes.

### IDENTIFICATION OF CODING SEQUENCES

Coding sequences from the surrounding DNA are identified, and these sequences are screened until a probable candidate cDNA is found. Much of the human genome will be sequenced over the next few years, in which case it may  
5 become feasible to identify coding sequences through database screening.

Candidates may also be identified by scanning databases consisting of partially sequenced cDNAs (Adams et al., 1991), known as expressed sequence tags, or ESTs. These resources are already largely developed, and include upwards of 100,000 cDNAs, the majority expressed primarily in the brain. It is not yet clear,  
10 however, that the complete set of cDNAs will be mapped to specific chromosomal locations in the near future, and that their data will soon be made publicly available. The database can be used to identify all cDNAs that map to the minimal candidate region for BP-I. These cDNAs are then used as probes to hybridize to the P1 contig, and new microsatellites are isolated, which are used to  
15 genotype the "LD" sample. Maximal linkage disequilibrium in the vicinity of one or two cDNAs is identified. These cDNAs are the first ones used to screen patient DNA for mutations. Database screening has already been used to identify a gene responsible for familial colon cancer (Papadopolous et al., 1993).

Coding sequences are also identified by exon amplification (Duyk et al.,  
20 1990; Buckler et al., 1991). Exon amplification targets exons in genomic DNA by identifying the consensus splice sequences that flank exon-intron boundaries. Briefly, exons are trapped in the process of cloning genomic DNA (e.g. from P1s) into an expression vector (Zhang et al., 1994). These clones are transfected into COS cells, RT-PCR is performed on total or cytoplasmic RNA isolated from the  
25 COS cells using primers that are complementary to the splicing vector. Exon amplification is tedious but routine; for example, the system developed by Buckler et al. (1991). This method is probably preferable to another widely used approach, direct selection, which involves screening cDNAs using large insert clone contigs, with several steps to maximize the efficiency of hybridization and  
30 recovery of the appropriate hybrid (Lovett et al., 1991). Although direct selection is more efficient than exon amplification (Del Mastro et al., 1994), it may not be practical as it depends on the candidate cDNA being expressed in the tissue from

which the cDNA library was made; there is no prior information to indicate the tissue or developmental stage in which BP-I genes would be expressed.

Once cDNAs are identified the most plausible candidates are screened by direct sequencing, SSCP or using chemical cleavage assays (Cotton et al. 1988).

5       The data are also evaluated for clues to the possible identity or mode of action of BP-I mutations. For example, it is known that trinucleotide repeat expansion is associated with the phenomenon of anticipation, or the tendency for a phenotype to become more severe and display an earlier age of onset in the lower generations of a pedigree (Ballabio, 1993). Several investigators have suggested  
10       that segregation patterns of BP-I are consistent with anticipation (McInnis et al., 1993; Nylander et al., 1994). The apparent transmission of BP-I, in association with the conserved 18q23 haplotype is constant with anticipation. Therefore, once the candidate region is narrowed to its minimal extent, the P1 clones are screened using trinucleotide repeat oligonucleotides (Hummerich et al., 1994). A PCR  
15       assay is developed and patient DNAs are screened for expanded alleles.

Genetic and physical data help to map the bipolar mood disorder gene to the 5 cM 18pter region of chromosome 18. New markers from this region are tested in order to locate the bipolar mood disorder gene in a region small enough to provide higher quality genetic tests for bipolar mood disorder, and to  
20       specifically find the mutated gene. Narrowing down the region in which the gene is located will lead to sequencing of the bipolar mood disorder gene as well as cloning thereof. Further genetic analysis employing, for example, new polymorphisms flanking D18S59 and D18S476 as well as the use of cosmids, yeast artificial chromosome (YAC) clones, or mixtures thereof, are employed in the  
25       narrowing down process. The next step in narrowing down the candidate region includes cloning of the chromosomal region 18pter including proximal and distal markers in a contig formed by overlapping cosmids and YACS. Subsequent subcloning in cosmids, plasmids or phages will generate additional probes for more detailed mapping.

30       The next step of cloning the gene involves exon trapping, screening of cDNA libraries, Northern blots or rt PCR (reverse transcriptase PCR) of samples from affected and unaffected individuals, direct sequencing of exons or testing

exons by SSCP (single strand conformation polymorphism), RNase protection or chemical cleavage.

Flanking markers on both sides of the bipolar mood disorder gene combined with D18S59 and D18S476 or a number of well-positioned markers that cover the chromosomal region (5 cM 18pter) carrying the disease gene, can give a high probability of affected or non-affected chromosomes in the range of 80-90% accuracy, depending on the informativeness of the markers used and their distance from the disease gene. Using current markers linked to bipolar mood disorder, and assuming closer flanking markers will be identified, a genetic test for families with bipolar mood disorder will be for diagnosis in conjunction with clinical evaluation, screening of risk and carrier testing in healthy siblings. In the future, subsequent delineation of closely linked markers which may show strong disequilibrium with the disorder, or identification of the defective gene, could allow screening of the entire at-risk population to identify carriers, and provide improved treatments.

#### TREATMENT OF BP-I PATIENTS USING GENOTYPE DATA

Using the fine mapping techniques described herein, BP-I susceptibility loci or genes in the 5 cM 18pter region in particular in the region #1 between SAVA5 and ga203, are identified and used to genotype patients diagnosed phenotypically with BP-I. Preferably, markers in the roughly 500 kb region defined by SAVA5 and ga203, inclusive, are used. More preferably, markers in either the region defined by D18S59 and w3422, inclusive, are used.

Genotyping with the markers described herein as well as additional markers permits confirmation of phenotypic BP-I diagnoses or assist with ambiguous clinical phenotypes which make it difficult to distinguish between BP-I and other possible psychiatric illnesses. A patient's genotype in the 5 cM 18pter region is determined and compared with previously determined genotypes of other individuals previously diagnosed with BP-I. Once an individual is genotyped as having a BP-I susceptibility locus in the 5 cM 18pter region, the individual is treated with any of the known methods effective in treating at least certain

individuals affected with BP-I, such as the administration of lithium salts, carbamazepine or valproic acid.

Studies are conducted correlating effective treatments with BP-I genotypes in the 5 cM 18pter region to determine the most effective treatments for particular  
5 genotypes. BP-I patients can then be genotyped in the 5 cM 18pter region and the statistically most effective treatment can be determined as a first course of therapy.

All publications and patent applications mentioned in this specification are herein incorporated by reference to the same extent as if each individual publication or patent application was specifically and individually indicated to be  
10 incorporated by reference.

The invention now being fully described, it will be apparent to one of ordinary skill in the art that many changes and modifications can be made thereto without departing from the spirit or scope of the appended claims.

## Appendix A

Consider the original mutation to have occurred on a chromosomal segment between two markers A and B. At the time the mutation was introduced, there were  $n_a$  alleles at marker A and  $n_b$  alleles at marker B. On the chromosome containing the disease mutation both marker A and marker B carried allele X. The probability that after  $g$  generations an affected individual carrying the original disease mutation would still have allele X at markers A and B is:

$$(1-\theta_1)^g(1-\theta_2)^g + (1-\theta_1)^g(1-(1-\theta_2)^g)f(X_B) + (1-(1-\theta_1)^g)(1-\theta_2)^gf(X_A) + (1-(1-\theta_1)^g)(1-(1-\theta_2)^g)f(X_A)f(X_B)$$

eq (1)

where  $\theta_1$  is the recombination fraction between disease and marker A,  $\theta_2$  is the recombination fraction between disease and marker B,  $g$  is the number of generations since founding (i.e. since the mutation was introduced into the population),  $f(X_A)$  is the population frequency of the X-allele at marker A and  $f(X_B)$  is the population frequency of the X-allele at marker B. This equation includes terms for the possibility of recombination between the markers and the disease locus, with the X-allele at the markers then being identical by state (IBS) rather than IBD. The probabilities of an affected individual with the original mutation having other haplotypes can be formulated similarly. The probability of having allele Z at marker B (where Z is any allele at marker B besides X) would be:

$$(1-\theta_1)^g(1-(1-\theta_2)^g)f(Z_B) + (1-(1-\theta_1)^g)(1-(1-\theta_2)^g)f(X_A)f(Z_B)$$

eq (2)

where  $f(Z_B)$  is the frequency of allele Z at marker B in the population. The probability of having allele Z at marker A (where Z is any allele at marker B besides X) would be :

$$(1-\theta_2)^g(1-(1-\theta_1)^g)f(Z_A) + (1-(1-\theta_1)^g)(1-(1-\theta_2)^g)f(X_B)f(Z_A)$$

eq (3)

where  $f(Z_A)$  is the frequency of allele Z at marker A in the population. Finally, the probability of having allele Z at both markers A and B would be:

$$(1-(1-\theta_1)^g)(1-(1-\theta_2)^g)f(Z_A)f(Z_B)$$

eq (4)

These probabilities assume (1) no interference in recombination and (2) the same marker alleles are present now as were present  $g$  generations ago, in similar frequencies. If, for example, marker A has  $n_a$  alleles and marker B has  $n_b$  alleles, then these probabilities form a  $(n_a) \times (n_b)$  by  $(n_a) \times (n_b)$  transition matrix, with row  $i$  containing the probabilities that founder haplotype  $i$  gave rise to each of the  $(n_a) \times (n_b)$  different haplotypes in  $g$  generations. The rows of this transition matrix sum to 1.

In simulations, the haplotype frequencies in the disease population were formulated using these transition probabilities, assuming the disease arose on a haplotype with the "1" allele at each of the two markers.

Once these transition probabilities are estimated, the likelihood of a particular founder chromosome giving rise to the observed sample of disease haplotypes in  $g$  generations is easily estimated. For example, if one assumed that the disease mutation arose on a chromosome with the X-allele at both markers, the likelihood ( $L_{X-X}$ ) that this chromosome was the founder of the present-day sampled disease chromosomes is given by the multinomial:

$$L_{X-X} = \prod_{i=1}^K (p_{X-X,i})^{Y_i}$$

eq (5)

where  $i$  indexes the  $K$  potential haplotypes for the two markers ( $K = (n_a)(n_b)$ ),  $p_{X-X,i}$  is the probability that the ancestral disease chromosome with the X-allele at both markers gave rise to a haplotype of type  $i$  in  $g$  generations (taken from the transition matrix), and  $Y_i$  is the observed number of haplotype  $i$  in the sample ( $\sum_i(Y_i)$  = the number of chromosomes in the sample to be analyzed). The likelihood in eq (5) assumes that all affected individuals are independent. While, after many generations of separation from a common ancestor one might consider these

individuals to be independent, they are in fact related through a complex and unknown pedigree. The simplification of considering individuals as independent makes the likelihood much more tractable to compute.

The  $K$  likelihoods are then summed, and weighted by the probability of observing that particular haplotype in the population to produce an overall likelihood for the 2-marker segment:

$$L = \sum_{i=1}^K f_i L_i$$

eq (6)

where  $f_i$  is the frequency of haplotype  $i$  in the population. This overall likelihood calculation parallels the approach taken by Terwilliger (1995, eq (2)). The haplotype frequencies are estimated from the sample of normal chromosomes. In the event that the haplotype resulting in the largest contribution to the overall likelihood in eq (6) is not observed in the normal sample, the upper 95% confidence interval for this frequency is used, and the remaining haplotype frequencies rescaled accordingly.

This overall likelihood is compared to the null likelihood, which is generated in exactly the same manner, except that it is assumed the markers were unlinked to

the disease locus ( $\theta_1 = \theta_2 = 0.5$  in, for example, eqs (1-4)). The  $\log_{10}$  of this likelihood ratio is a LOD score. One might consider to use in the null likelihood transition probabilities calculated under the assumption of linkage equilibrium. Under this null the cells of the transition matrix are computed by multiplication of allele frequencies, assuming independence of marker loci. These two forms of the null likelihood are equivalent in value for  $g$  of approximately 20 or greater, and for  $g < 20$  the values are nearly equivalent.

Because  $\theta_1$  and  $\theta_2$  are obviously unknown, the putative disease locus is set to be in the middle of the segment and therefore  $\theta_1$  and  $\theta_2$  are one-half the genetic distance (converted to recombination fraction by the Haldane mapping function, (Ott 1991)) between the two marker loci forming the segment. In fact, one could estimate  $\theta_1$  and  $\theta_2$ , or their ratio, and the method could easily be modified to do so, however for our purposes finding a linked segment is suitable.

This basic procedure has been modified to deal with heterogeneity in the sample of disease chromosomes. Not all chromosomes in the disease sample may be true disease chromosomes from a common founder. Individuals heterozygous for the disease mutation will add one chromosome to the disease sample that will not be a true disease chromosome. Additionally, affected individuals not linked to the

particular chromosomal segment being analyzed (either because they are phenocopies or because of locus heterogeneity) will contribute two chromosomes to the affected sample that do not harbor this disease locus. When the null hypothesis of no linkage is not true, some fraction,  $\alpha$ , of the chromosomes in the disease sample will be associated with this chromosomal segment, and  $(1-\alpha)$  will not be associated. We decided to examine  $\alpha$  in steps of 0.1, from 1.0 to 0.0, and for each step in  $\alpha$  produce a new transition matrix under the alternative hypothesis and calculate a LOD score. If we call the transition matrix calculated under the alternative hypothesis (where the disease locus is hypothesized to be in the middle of the 2-marker segment)  $T_a$  and call the transition matrix calculated under the null hypothesis (where the disease locus is unlinked to the 2-marker segment)  $T_n$ , then a new transition matrix for the alternative hypothesis is calculated as:

$$T^*_a = \alpha T_a + (1 - \alpha) T_n$$

eq (7)

The transition matrix under the null uses  $\alpha=0$ . The LOD score is then maximized over the one parameter  $\alpha$ .

WHAT IS CLAIMED IS:

1. A method of detecting the presence of a bipolar mood disorder susceptibility locus in an individual comprising:
  - 5 analyzing a sample of DNA from said individual for the presence of a DNA polymorphism on the short arm of chromosome 18 between SAVA5 and ga203, wherein said DNA polymorphism is associated with a form of bipolar mood disorder.
  - 10 2. The method of claim 1, wherein said DNA polymorphism is located on the short arm of chromosome 18 between D18S1140 and ga203, inclusive.
  3. The method of claim 1, wherein said DNA polymorphism is located on the short arm of chromosome 18 between SAVA5 and W3422, inclusive.
  - 15 4. The method of claim 1, wherein said DNA polymorphism is located on the short arm of chromosome 18 between D18S1140 and W3422, inclusive.
  5. The method of claim 1, wherein said DNA polymorphism is located on the short arm of chromosome 18 between D18S1140 and at201, inclusive.
  - 20 6. The method of claim 1, wherein said DNA polymorphism is located on the short arm of chromosome 18 between D18S1140 and ta201, inclusive.
  - 25 7. The method of claim 1, wherein said DNA polymorphism is located on the short arm of chromosome 18 between D18S59 and ta201, inclusive.

8. The method of claim 1, wherein said analyzing further comprises:
- a. obtaining DNA samples from family members of said individual,
  - b. analyzing said DNA samples from family members for the presence of said DNA polymorphism, and
  - 5 c. correlating the presence or absence of the DNA polymorphism with a phenotypic diagnosis of bipolar mood disorder for said individual and for said family members.
9. A method for detecting the presence of a DNA polymorphism linked to a
- 10 gene associated with bipolar mood disorder in an individual comprising:
- a. typing blood relatives of said individual for a DNA polymorphism located within a 500kb region of chromosome 18, wherein said region is located between SAVA5 and ga203, inclusive,
  - b. analyzing a DNA sample from said individual for the presence of
  - 15 said DNA polymorphism.
10. A method of genetically diagnosing bipolar mood disorder in an individual comprising:
- a. obtaining a DNA sample from said individual,
  - 20 b. analyzing said DNA sample for the presence of a DNA polymorphism associated with bipolar mood disorder, wherein said DNA polymorphism is located within a 500 kb region of chromosome 18, wherein said region is located between SAVA5 and ga203, inclusive.
- 25 11. A method of confirming a phenotypic diagnosis of bipolar mood disorder in an individual comprising:
- a. obtaining a DNA sample from said individual,
  - b. analyzing said DNA sample for the presence of a DNA
  - 30 polymorphism associated with bipolar mood disorder, wherein said DNA polymorphism is located within a 500 kb region of chromosome 18, wherein said region is located between SAVA5 and ga203, inclusive.

12. The method of claim 10, wherein said individual has Spanish or Amerindian ancestry.
13. A method of classifying subtypes of bipolar mood disorder comprising:
- 5 a. identifying one or more DNA polymorphisms located within a 500 kb region of chromosome 18, wherein said region is located between SAVA5 and ga203, inclusive; and
- b. analyzing DNA samples from individuals phenotypically diagnosed with bipolar mood disorder for the presence or absence of one of more of said
- 10 DNA polymorphisms.
14. A method of treating an individual diagnosed with bipolar mood disorder comprising:
- a. identifying one or more DNA polymorphisms located within a 500
- 15 kb region of chromosome 18, wherein said region is located between SAVA5 and ga203, inclusive; and
- b. analyzing DNA samples from individuals phenotypically diagnosed with bipolar mood disorder for the presence or absence of one of more of said DNA polymorphisms, and
- 20 c. selecting a treatment plan that is most effective for individuals having a particular genotype within said 500 kb region of chromosome 18.
15. An isolated polynucleotide capable of selectively hybridizing with a DNA sample from an individual phenotypically diagnosed with severe bipolar mood
- 25 disorder, wherein said polynucleotide does not selectively hybridize with a DNA sample from an individual not affected by severe bipolar mood disorder, wherein said isolated polynucleotide selectively hybridizes with a complementary polynucleotide within a 500 kb region of chromosome 18, wherein said region is located between SAVA5 and ga203, inclusive.

30

16. The isolated polynucleotide of claim 15, wherein said complementary polynucleotide is within a 500 kb region of chromosome 18, between SAVA5 and ga203, inclusive.

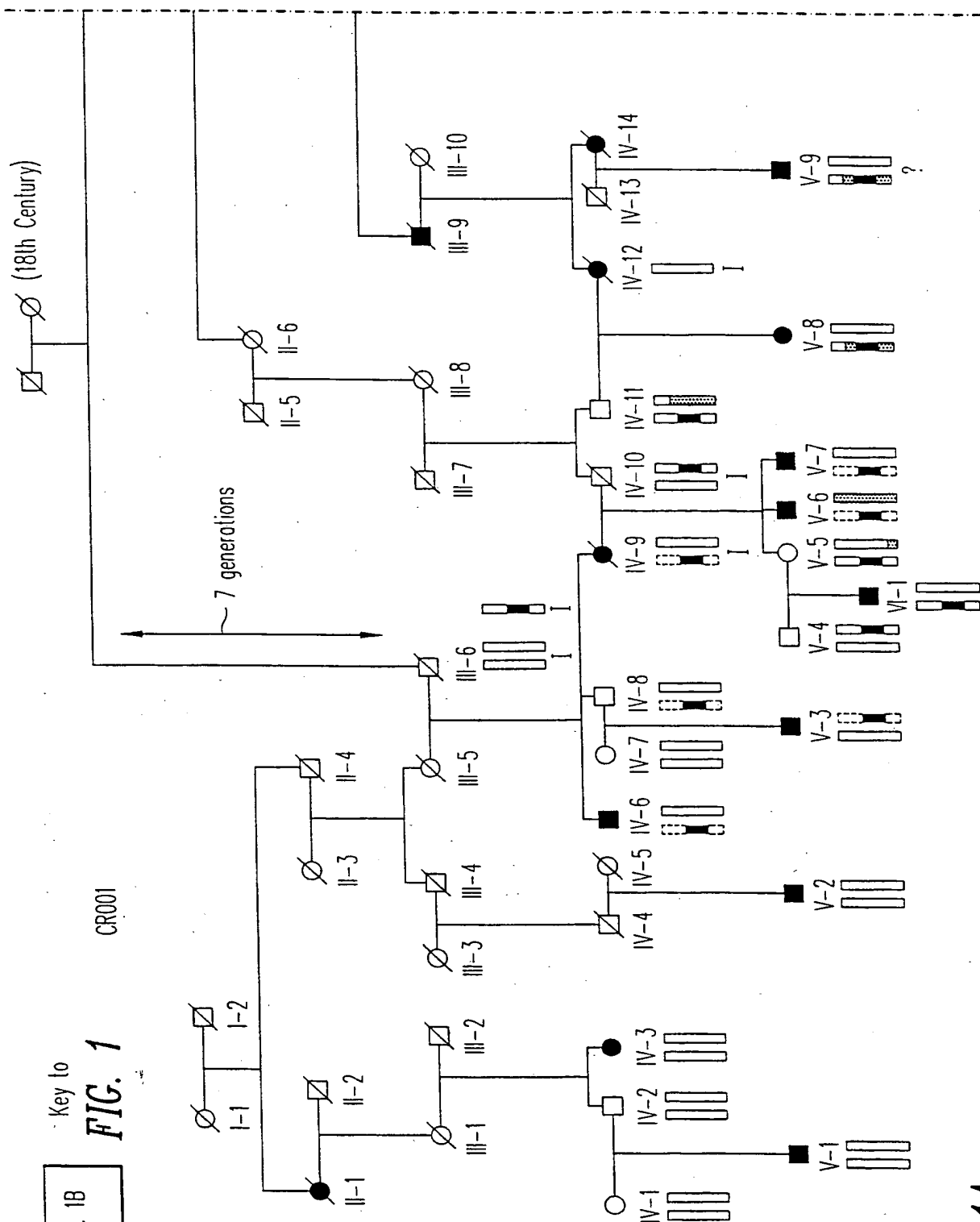


FIG. 1A

2/57

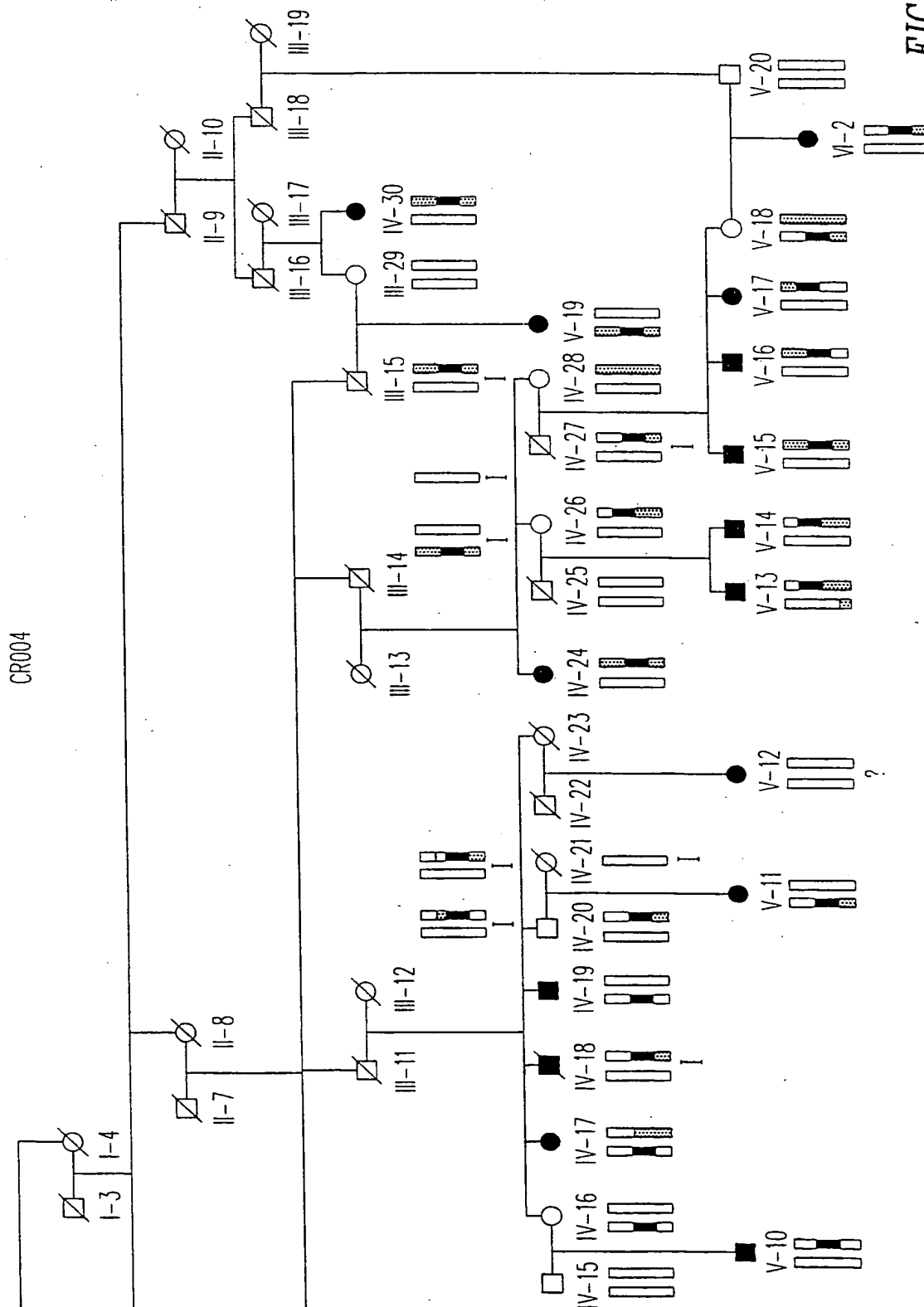


FIG. 1B

3/57

		Family CR001		Family CR004		Combined	
Marker Name	distance from pter	Z <sub>max</sub> ≥ 0.8	Theta	Z <sub>max</sub> ≥ 1.2	Theta	Z <sub>max</sub> ≥ 1.6	Theta
D1S456	224.6	<b>1.32</b>	0.0	0.0	0.50	0.0	0.50
D2S130	230.1	<b>0.89</b>	0.0	0.12	0.35	0.36	0.26
D3S1285	91.0	0.00	0.50	<b>2.59</b>	0.00	1.15	0.16
D4S171	207.9	<b>1.07</b>	0.07	0.01	0.05	0.22	0.29
D5S427	69.6	<b>1.39</b>	0.0	0.0	0.50	0.7	0.18
D7S510	60.5	0.04	0.40	<b>2.04</b>	0.0	0.82	0.17
D11S929	36.3	<b>0.80</b>	0.11	0.03	0.42	0.43	0.24
D11S1392	38.6	<b>0.86</b>	0.07	0.90	0.23	<b>1.58</b>	0.19
D11S1312	42.0	0.47	0.13	<b>1.77</b>	0.0	<b>1.95</b>	0.05
D13S175	7.4	<b>0.83</b>	0.0	0.0	0.50	0.24	0.15
D15S126	45.5	<b>1.09</b>	0.0	0.0	0.48	0.06	0.40
D16S521	4.6	<b>1.46</b>	0.0	0.41	0.26	1.18	0.17
D16S515	94.8	<b>0.93</b>	0.09	0.01	0.46	0.39	0.25
D16S486	133.6	0.27	0.19	<b>1.29</b>	0.20	<b>1.60</b>	0.20
D17S849	0.60	0.0	0.50	<b>1.22</b>	0.07	0.32	0.14
D18S59	1.1	<b>1.43</b>	0.0	0.0	0.50	0.02	0.46
D18S1105	2.8	<b>0.97</b>	0.0	0.01	0.47	0.01	0.46
D18S71	43.8	<b>0.96</b>	0.0	0.0	0.50	0.0	0.50
D18S64	84.0	0.33	0.11	<b>1.34</b>	0.15	<b>1.67</b>	0.13
D18S55	95.5	0.0	0.50	<b>2.09</b>	0.13	1.51	0.18
D18S61	103.8	0.0	0.50	<b>2.26</b>	0.12	<b>1.94</b>	0.16
D18S488	105.6	0.0	0.50	<b>1.26</b>	0.14	1.02	0.19
D18S1161	113.0	0.0	0.50	<b>1.79</b>	0.16	<b>1.76</b>	0.17

FIG. 2

4/57

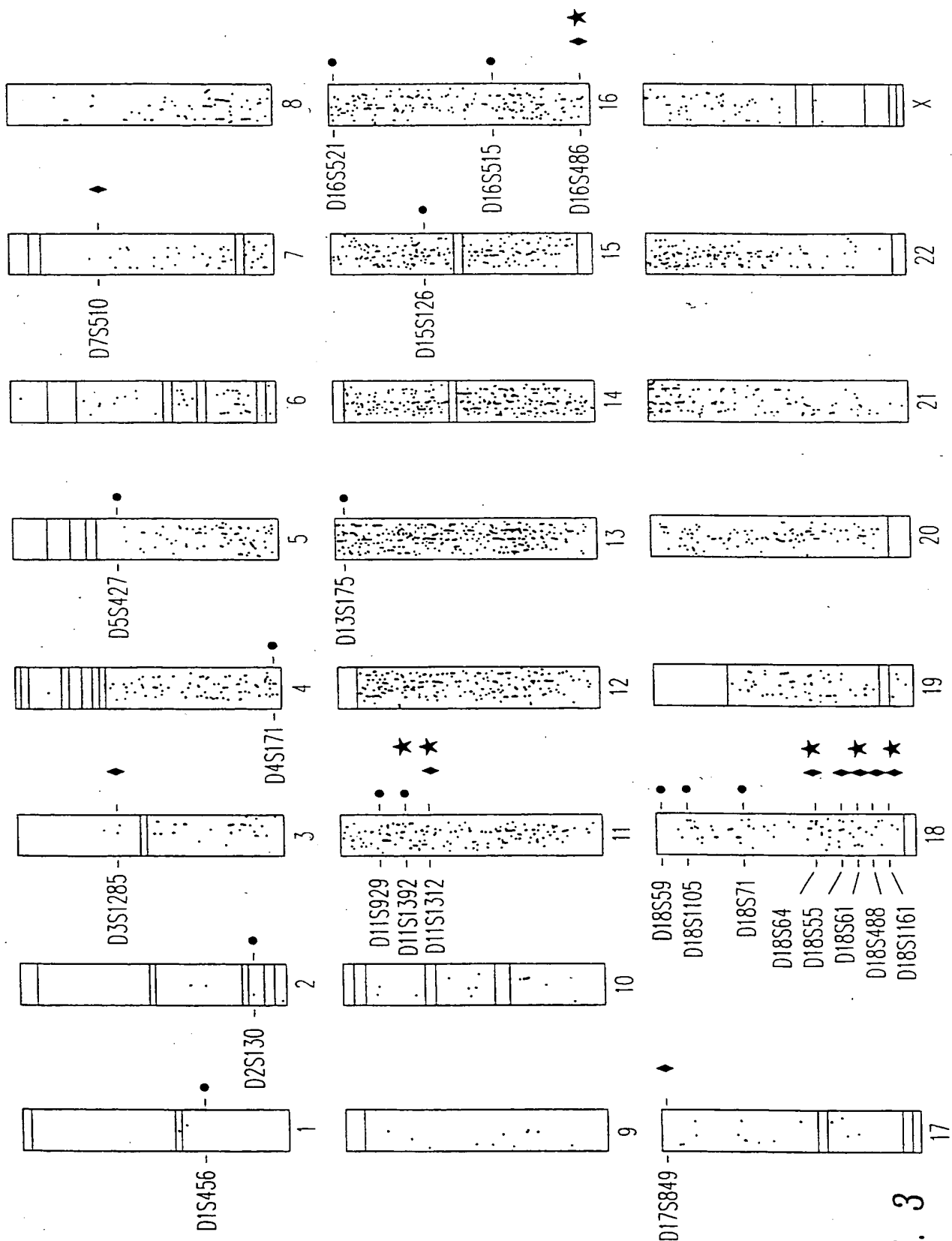


FIG. 3

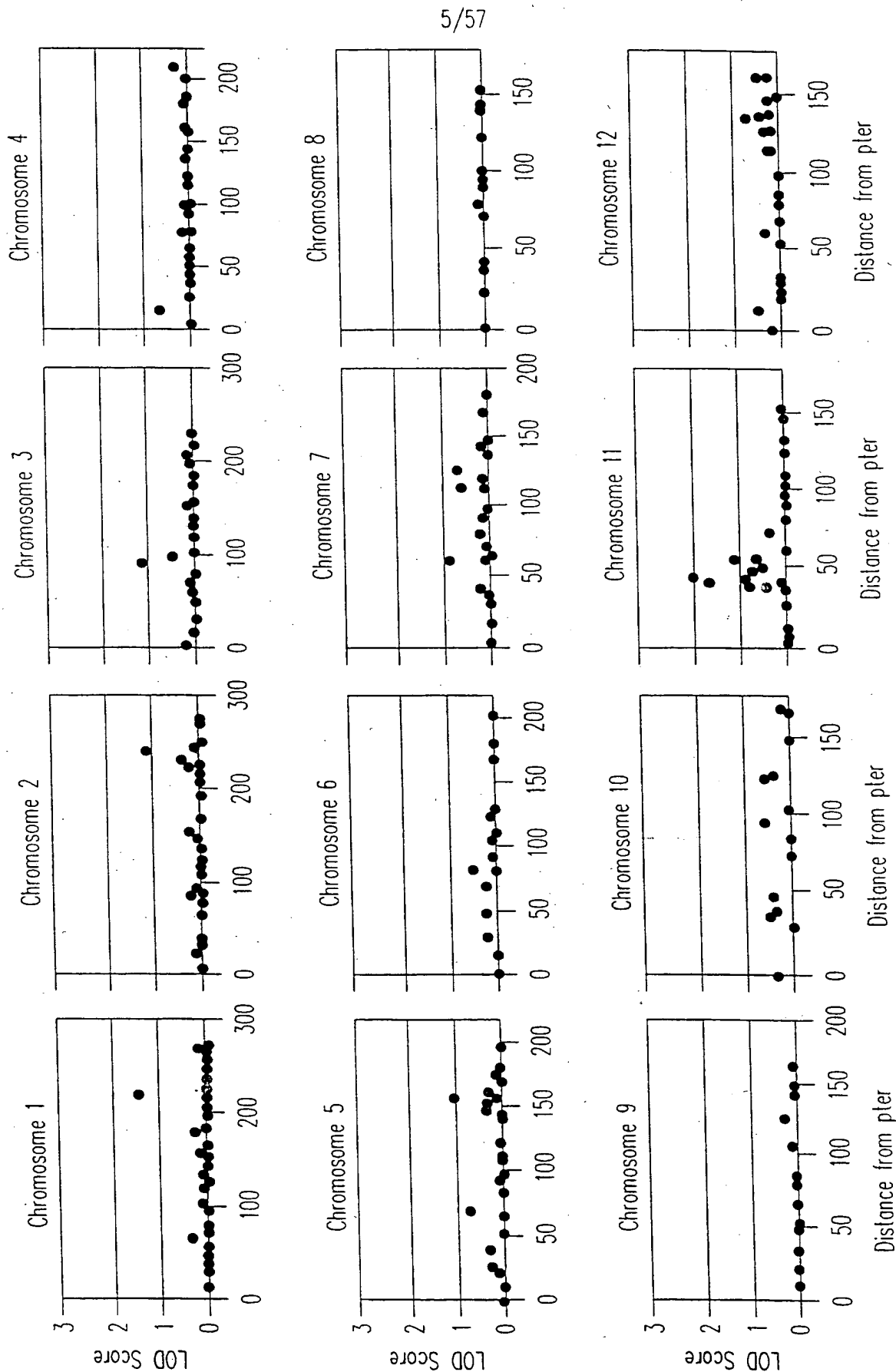


FIG. 4A

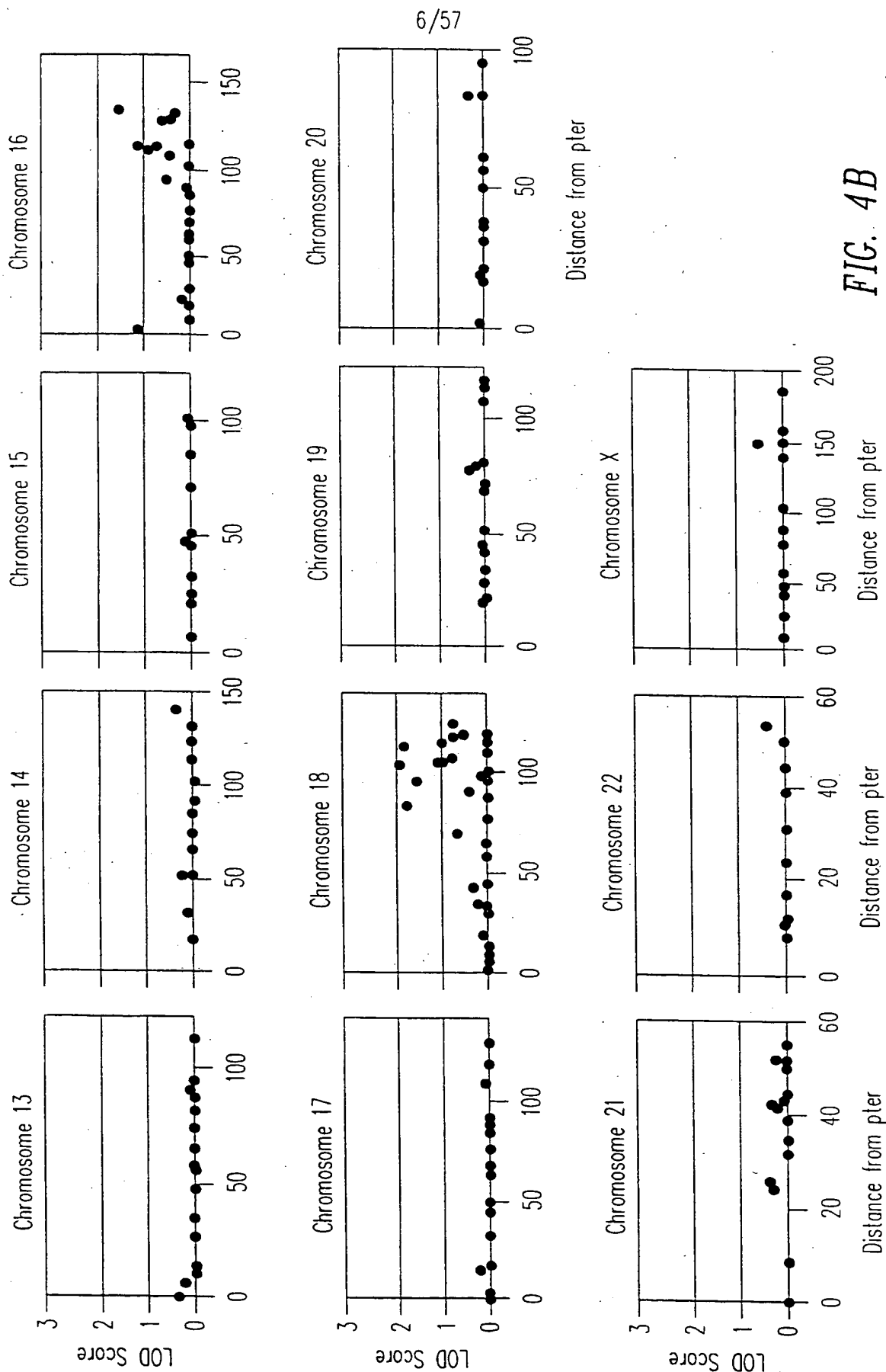


FIG. 4B

7/57

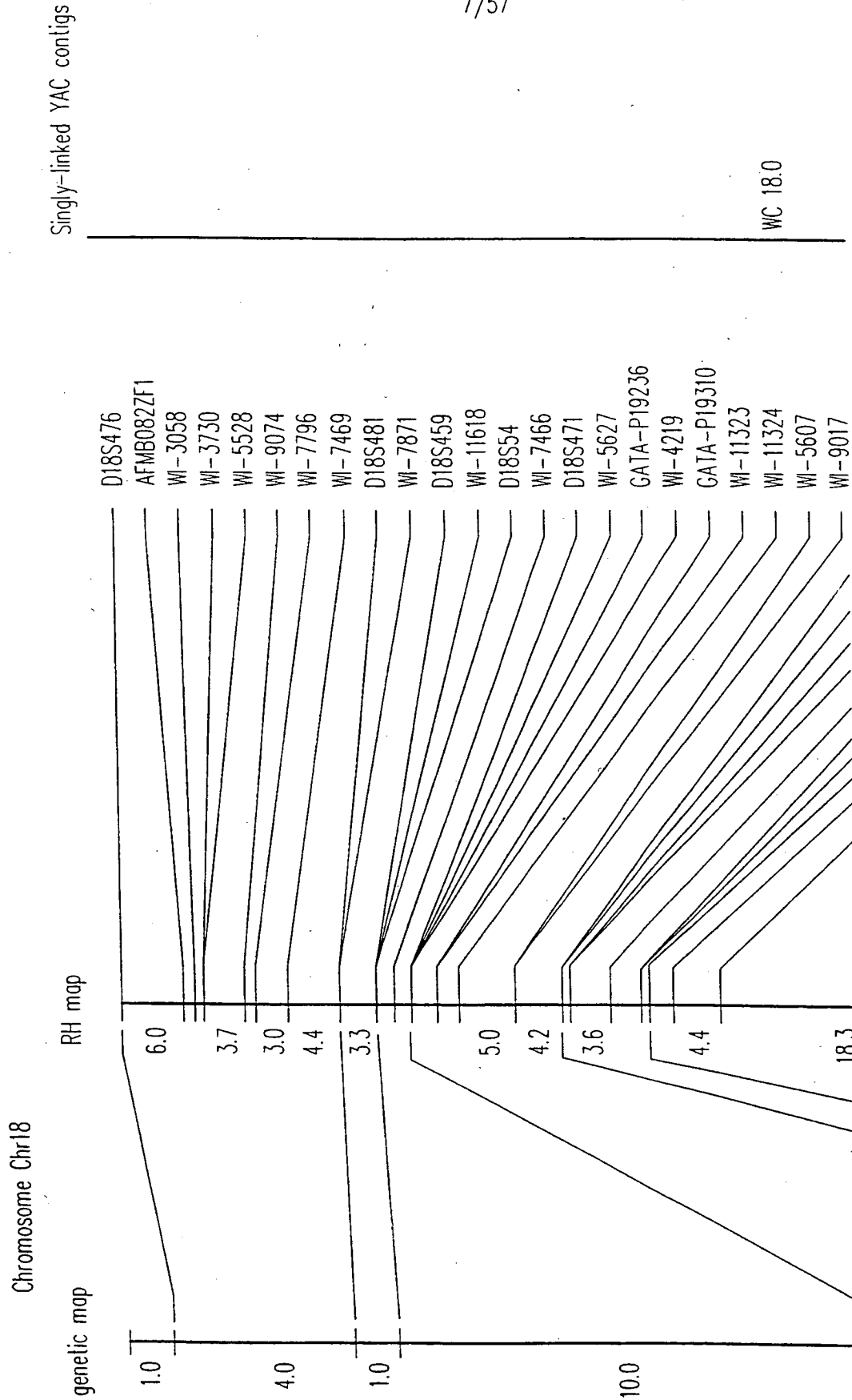


FIG. 5

8/57

	STS	Chrom	Map Position		Contig	
			Genetic	RH	Single	Double
1	<u>WI-9527</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1465</u>
2	<u>CHLC.GGAT2G04</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1465</u>
3	<u>CHLC.GGAT2G04.1217</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1465</u>
4	<u>D18S59</u>	<u>Chr18</u>	<u>0 cM</u>	-	<u>WC18.0</u>	<u>WC-1465</u>
5	<u>D18S1140</u>	<u>Chr18</u>	<u>0 cM</u>	-	<u>WC18.0</u>	<u>WC-1465</u>
6	<u>WI-7796</u>	<u>Chr18</u>	-	<u>15 cR</u>	<u>WC18.0</u>	-
7	<u>WI-9074</u>	<u>Chr18</u>	-	<u>12 cR</u>	<u>WC18.0</u>	<u>WC-1465</u>
8	<u>WI-5528</u>	<u>Chr18</u>	-	<u>7 cR</u>	<u>WC18.0</u>	-
9	<u>D18S476</u>	<u>Chr18</u>	<u>1 cM</u>	<u>0 cR</u>	<u>WC18.0</u>	-
10	<u>WI-7226</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-909</u>
11	<u>AFMB324ZE5</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-909</u>
12	<u>AFMB082ZF1</u>	<u>Chr18</u>	-	<u>5 cR</u>	<u>WC18.0</u>	<u>WC-909</u>
13	<u>D18S1146</u>	<u>Chr18</u>	<u>1 cM</u>	-	<u>WC18.0</u>	<u>WC-909</u>
14	<u>WI-3058</u>	<u>Chr18</u>	-	<u>5 cR</u>	<u>WC18.0</u>	<u>WC-909</u>
15	<u>D18S1105</u>	<u>Chr18</u>	<u>1 cM</u>	-	<u>WC18.0</u>	<u>WC-909</u>
16	<u>WI-3730</u>	<u>Chr18</u>	-	<u>5 cR</u>	<u>WC18.0</u>	<u>WC-1576</u>
17	<u>AFM077YD11</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1576</u>
18	<u>D18S1098</u>	<u>Chr18</u>	<u>4 cM</u>	-	<u>WC18.0</u>	<u>WC-1576</u>
19	<u>WI-7469</u>	<u>Chr18</u>	-	<u>16 cR</u>	<u>WC18.0</u>	<u>WC-1576</u>
20	<u>WI-7871</u>	<u>Chr18</u>	-	<u>22 cR</u>	<u>WC18.0</u>	<u>WC-1576</u>
21	<u>D18S481</u>	<u>Chr18</u>	<u>5 cM</u>	<u>21 cR</u>	<u>WC18.0</u>	<u>WC-1576</u>
22	<u>WI-4747</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1576</u>
23	<u>D18S1154</u>	<u>Chr18</u>	<u>6 cM</u>	-	<u>WC18.0</u>	<u>WC-1576</u>
24	<u>CHLC.ATA14B09</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1576</u>
25	<u>WI-7466</u>	<u>Chr18</u>	-	<u>27 cR</u>	<u>WC18.0</u>	<u>WC-1576</u>
26	<u>D18S54</u>	<u>Chr18</u>	<u>6 cM</u>	<u>25 cR</u>	<u>WC18.0</u>	<u>WC-1576</u>
27	<u>D18S63</u>	<u>Chr18</u>	<u>6 cM</u>	-	<u>WC18.0</u>	<u>WC-1576</u>
28	<u>D18S459</u>	<u>Chr18</u>	<u>6 cM</u>	<u>25 cR</u>	<u>WC18.0</u>	<u>WC-1576</u>
29	<u>WI-6014</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1576</u>
30	<u>WI-4219</u>	<u>Chr18</u>	-	<u>28 cR</u>	<u>WC18.0</u>	<u>WC-143</u>
31	<u>AFM238YG3</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-143</u>
32	<u>D18S471</u>	<u>Chr18</u>	<u>17 cM</u>	<u>28 cR</u>	<u>WC18.0</u>	<u>WC-143</u>
33	<u>D18S458</u>	<u>Chr18</u>	<u>17 cM</u>	-	<u>WC18.0</u>	<u>WC-143</u>

FIG. 6A

9/57

34	<u>D18S452</u>	<u>Chr18</u>	<u>17 cM</u>	-	<u>WC18.0</u>	<u>WC-143</u>
35	<u>D18S62</u>	<u>Chr18</u>	<u>17 cM</u>	-	<u>WC18.0</u>	<u>WC-143</u>
36	<u>WI-5627</u>	<u>Chr18</u>	-	<u>28 cR</u>	<u>WC18.0</u>	<u>WC-143</u>
37	<u>CHLC.GATA82D03</u>	<u>Chr18</u>	-	<u>28 cR</u>	<u>WC18.0</u>	<u>WC-143</u>
38	<u>FB25F12</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-143</u>
39	<u>CHLC.GATA51H07</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-143</u>
40	<u>CHLC.GATA88A12</u>	<u>Chr18</u>	-	<u>30 cR</u>	<u>WC18.0</u>	<u>WC-143</u>
41	<u>WI-9619</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-143</u>
42	<u>AFMB346YA9</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-143</u>
43	<u>AFM323TC9</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-862</u>
44	<u>WI-5607</u>	<u>Chr18</u>	-	<u>36 cR</u>	<u>WC18.0</u>	<u>WC-862</u>
45	<u>WI-9017</u>	<u>Chr18</u>	-	<u>36 cR</u>	<u>WC18.0</u>	<u>WC-862</u>
46	<u>AFM077YF7</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-934</u>
47	<u>WI-8546</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-934</u>
48	<u>CHLC.GGAA16G02</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-934</u>
49	<u>D18S464</u>	<u>Chr18</u>	<u>32 cM</u>	<u>46 cR</u>	<u>WC18.0</u>	<u>WC-934</u>
50	<u>NIB1802</u>	<u>Chr18</u>	-	<u>56 cR</u>	<u>WC18.0</u>	<u>WC-934</u>
51	<u>D18S1153</u>	<u>Chr18</u>	<u>34 cM</u>	-	<u>WC18.0</u>	<u>WC-934</u>
52	<u>D18S1150</u>	<u>Chr18</u>	<u>36 cM</u>	-	<u>WC18.0</u>	<u>WC-934</u>
53	<u>WI-4589</u>	<u>Chr18</u>	-	<u>58 cR</u>	<u>WC18.0</u>	<u>WC-934</u>
54	<u>WI-4319</u>	<u>Chr18</u>	-	<u>62 cR</u>	<u>WC18.0</u>	<u>WC-934</u>
55	<u>D18S1158</u>	<u>Chr18</u>	<u>38 cM</u>	-	<u>WC18.0</u>	<u>WC-934</u>
56	<u>D18S1116</u>	<u>Chr18</u>	<u>40 cM</u>	-	<u>WC18.0</u>	<u>WC-377</u>
57	<u>CHLC.GATA11A06.668</u>	<u>Chr18</u>	-	<u>48 cR</u>	<u>WC18.0</u>	<u>WC-377</u>
58	<u>CHLC.GATA11A06</u>	<u>Chr18</u>	-	<u>54 cR</u>	<u>WC18.0</u>	<u>WC-377</u>
59	<u>D18S53</u>	<u>Chr18</u>	<u>41 cM</u>	-	<u>WC18.0</u>	<u>WC-377</u>
60	<u>WI-9134</u>	<u>Chr18</u>	-	<u>52 cR</u>	<u>WC18.0</u>	<u>WC-377</u>
61	<u>IB1114</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-377</u>
62	<u>D18S482</u>	<u>Chr18</u>	<u>41 cM</u>	<u>56 cR</u>	<u>WC18.0</u>	<u>WC-377</u>
63	<u>WI-2382</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-377</u>
64	<u>WI-6819</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-377</u>
65	<u>D18S71</u>	<u>Chr18</u>	<u>43 cM</u>	<u>84 cR</u>	<u>WC18.0</u>	<u>WC-377</u>
66	<u>AFMA058YG5</u>	<u>Chr18</u>	-	<u>80 cR</u>	<u>WC18.0</u>	<u>WC-377</u>
67	<u>WI-5506</u>	<u>Chr18</u>	-	<u>90 cR</u>	<u>WC18.0</u>	<u>WC-377</u>
68	<u>D18S453</u>	<u>Chr18</u>	<u>43 cM</u>	<u>93 cR</u>	<u>WC18.0</u>	<u>WC-738</u>
69	<u>D18S73</u>	<u>Chr18</u>	<u>43 cM</u>	-	<u>WC18.0</u>	<u>WC-377</u>
70	<u>STSG-10174</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-377</u>

FIG. 6B

10/57

71	<u>CHLC.GCT5D07</u>	<u>Chr18</u>	-	<u>101 cR</u>	<u>WC18.0</u>	<u>WC-377</u>
72	<u>WI-10768</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1182</u>
73	<u>D18S1149</u>	<u>Chr18</u>	<u>49 cM</u>	-	<u>WC18.0</u>	<u>WC-1182</u>
74	<u>WI-1869</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1182</u>
75	<u>D18S1104</u>	<u>Chr18</u>	<u>49 cM</u>	-	<u>WC18.0</u>	<u>WC-1182</u>
76	<u>AFMA205YH5</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1182</u>
77	<u>AFMB340VE5</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1182</u>
78	<u>CHLC.GATA41C05</u>	<u>Chr18</u>	-	<u>185 cR</u>	<u>WC18.0</u>	<u>WC-1182</u>
79	<u>AFMB319WF9</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1182</u>
80	<u>D18S44</u>	<u>Chr18</u>	-	-	<u>WC18.0</u>	<u>WC-1182</u>

Details on contig assembly.*FIG. 6C*

11/57

18p allele frequencies

MARKERNAME		off 105	ntrans	control		
D18SAVA5	225	0.04	0.02			
	227	0.29	0.24			
	229	0.22	0.15			
	231	0.04	0.08			
	233	0.14	0.23			
	235	0.25	0.22			
	237	0.02	0.03			
	239	0.00	0.00			
D18SCA211	183	0.02	0.04	0.01		
	189	0.00	0.01	0.01		
	191	0.01	0.00	0.03		
	193	0.24	0.17	0.33		
	195	0.21	0.19	0.18		
	197	0.06	0.11	0.03		
	199	0.06	0.04	0.01		
	201	0.10	0.14	0.10		
	203	0.02	0.04	0.06		
	205	0.16	0.18	0.14		
	207	0.09	0.04	0.06		
	209	0.02	0.02	0.02		
	211	0.01	0.00	0.00		
	215	0.00	0.00	0.00		
	217	0.00	0.00	0.01		
D18SCA212	200	0.40	0.40	0.39		
	202	0.31	0.32	0.29		
	204	0.05	0.05	0.03		
	206	0.04	0.06	0.10		
	214	0.01	0.00	0.00		
	216	0.14	0.12	0.15		

FIG. 7

12/57

18p allele frequencies

MARKERNAME		aff 105	ntrans	control		
	218	0.04	0.00	0.04		
D18S1140	256	0.06	0.07	0.06		
	268	0.77	0.72	0.73		
	270	0.02	0.00	0.06		
	272	0.03	0.03	0.03		
	274	0.00	0.00	0.00		
	276	0.03	0.06	0.02		
	278	0.02	0.06	0.05		
	280	0.04	0.06	0.02		
	282	0.01	0.00	0.02		
MARKERNAME		aff 105	ntrans	control		
D18S59	148	0.16	0.26	0.21		
	150	0.07	0.09	0.14		
	152	0.02	0.06	0.01		
	154	0.36	0.19	0.28	0.17	0.08
	156	0.04	0.04	0.08		
	158	0.22	0.21	0.13		
	160	0.04	0.08	0.05		
	162	0.05	0.06	0.05		
	164	0.02	0.01	0.02		
	168	0.00	0.00	0.01		
D18STA201	214	0.02	0.00	0.00		
	220	0.09	0.09	0.04		
	222	0.01	0.00	0.01		
	228	0.01	0.01	0.00		
	230	0.25	0.22	0.16	0.03	0.09

FIG. 7A

13/57

18p allele frequencies

[illegible]

FIG. 7B

**SUBSTITUTE SHEET (RULE 26)**

14/57

18p allele frequencies

MARKERNAME		aff 105	ntrans	control		
D18SAT201	170	0.53	0.55	0.52		
	174	0.00	0.01	0.01		
	<b>178</b>	0.37	0.36	0.36		
	182	0.01	0.00	0.00		
	186	0.07	0.06	0.07		
	190	0.01	0.00	0.00		
	194	0.01	0.01	0.03		
D18SCA225	160	0.16	0.20	0.21		
	168	0.02	0.04	0.00		
	170	0.00	0.00	0.01		
	<b>172</b>	<b>0.47</b>	<b>0.38</b>	<b>0.42</b>	<b>0.09</b>	<b>0.04</b>
	174	0.22	0.24	0.26		
	176	0.04	0.04	0.05		
	178	0.04	0.04	0.02		
	180	0.02	0.01	0.01		
	184	0.03	0.00	0.02		
D18SW3442	<b>10</b>	<b>0.42</b>	<b>0.28</b>	<b>0.36</b>	<b>0.14</b>	<b>0.06</b>
	12	0.01	0.01	0.01		
	14	0.07	0.11	0.11		
	16	0.12	0.17	0.12		
	18	0.18	0.15	0.14		
	20	0.05	0.09	0.09		
	22	0.08	0.10	0.11		
	24	0.05	0.08	0.03		
	26	0.00	0.00	0.02		
	38	0.00	0.00	0.00		
D18SCA213	112	0.12	0.17	0.07		
	120	0.00	0.05	0.01		
	122	0.03	0.03	0.04		
	<b>124</b>	<b>0.44</b>	<b>0.37</b>	<b>0.46</b>		

FIG. 7C

15/57

18p allele frequencies

MARKERNAME		off 105	ntrans	control		
	126	0.30	0.24	0.35		
	128	0.08	0.11	0.06		
	130	0.00	0.00	0.00		
	132	0.03	0.02	0.01		
D18SGAT201	142	0.04	0.06	0.02		
	146	0.08	0.08	0.06		
	150	0.61	0.62	0.69		
	154	0.15	0.15	0.12		
	158	0.11	0.07	0.10		
	162	0.02	0.02	0.00		
D18SGAT203						
	188	0.42	0.37	0.38		
	192	0.12	0.14	0.17		
	196	0.01	0.04	0.01		
	200	0.02	0.04	0.01		
	204	0.06	0.02	0.04		
	208	0.19	0.21	0.20		
	212	0.11	0.11	0.11		
	216	0.09	0.07	0.08		
D18SCA219	221	0.00		0.01		
	223	0.00		0.00		
	225	0.00		0.00		
	233	0.00		0.00		
	235	0.22		0.21		
	239	0.02		0.01		
	241	0.54		0.63		
	243	0.07		0.07		
	245	0.13		0.06		
MARKERNAME		off 105	ntrans	control		

FIG. 7D

16/57

18p allele frequencies

MARKERNAME		aff 105	ntrans	control		
D18S1105	101	0.16	0.11			
	103	0.12	0.08			
	105	0.03	0.02			
	81	0.02	0.01			
	83	0.01	0.02			
	85	0.51	0.54			
	87	0.01	0.06			
	91	0.00	0.00			
	95	0.01	0.04			
	97	0.04	0.04			
	99	0.08	0.06			
D18SCA209	173	0.57	0.53	0.69		
	175	0.02	0.03	0.04		
	<b>177</b>	<b>0.20</b>	0.18	<b>0.09</b>		
	179	0.01	0.03	0.00		
	181	0.19	0.24	0.18		
	187	0.00	0.00	0.00		
D18SCA202	182	0.16	0.14			
	184	0.02	0.00			
	186	0.01	0.01			
	190	0.09	0.02			
	192	0.10	0.16			
	194	0.10	0.09			
	196	0.37	0.35			
	198	0.09	0.10			
	200	0.05	0.08			
	202	0.00	0.03			
	208	0.00	0.00			
D18S1146	270	0.32	0.35			
	272	0.07	0.10			
	<b>274</b>	0.60	0.51			

FIG. 7E

17/57

18p allele frequencies

MARKERNAME		aff 105	ntrans	control		
	276	0.02	0.04			
<b>D18S166D05</b>	300	0.17	0.21	0.19		
	304	0.16	0.12	0.14		
	308	0.18	0.18	0.13		
	<b>312</b>	<b>0.35</b>	<b>0.26</b>	<b>0.36</b>	<b>* *</b>	
	316	0.08	0.18	0.11		
	320	0.04	0.04	0.03		
	324	0.01	0.01	0.02		
<b>D18S476</b>	261	0.00	0.01	0.01		
	263	0.01	0.04	0.04		
	265	0.05	0.12	0.04		
	267	0.20	0.26	0.23		
	269	0.08	0.09	0.04		
	<b>271</b>	<b>0.56</b>	<b>0.38</b>	<b>0.54</b>	<b>* * *</b>	
	273	0.04	0.08	0.07		
	275	0.04	0.03	0.03		

FIG. 7F

18/57

Affected haplotypes

18p	PAN	MAN	ca212		1140		59		ca231		ta201	
200	279	280	218	1	268	1	158	1	186	1	242	
200			200	1	268	1	158	1	186	1	248	
204	309	349	200	1	282	1	150	1	202	1	220	
204			206	1	268	1	158	1	184	1	250	
206	1	2	218	0	276	0	156	0	186	0	252	
206			200	0	268	0	148	0	184	0	248	
207	277	278	200	1	268	1	154	1	194	1	220	
207			204	1	268	1	158	1	184	1	230	
209	0	0	200	1	268	0	154	1	186	0	242	
209			200	1	256	0	150	1	184	0	254	
213	0	0	216	0	272	1	150	1	186	0	250	
213			200	0	282	1	150	1	184	0	238	
214	460	459	202	1	268	1	158	1	200	1	220	
214			216	1	276	1	154	1	186	1	242	
215	1	270	218	1	276	1	160	0	186	1	242	
215			200	1	268	1	154	0	186	1	230	
216	1	259	204	1	278	0	156	1	186	0	230	
216			200	1	268	0	162	1	184	0	252	
218	273	272	200	1	268	1	162	1	186	1	220	
218			200	1	268	1	158	1	186	1	246	
219	0	0	202	1	256	1	154	1	186	1	230	
219			200	1	268	1	168	1	184	1	250	
220	267	2	216	0	268	1	152	1	186	1	230	
220			200	0	268	1	154	1	186	1	232	
221	0	0	202	1	268	1	160	1	184	0	250	
221			202	1	268	1	154	1	186	0	250	
223	0	0	202	1	280	0	148	0	184	0	256	
223			202	1	268	0	154	0	186	0	252	
225	264	2	200	1	268	1	164	1	186	1	230	
225			200	1	268	1	158	1	186	1	246	
226	1	2	202	1	268	0	154	0	186	0	230	
226			202	1	256	0	148	0	184	0	254	
228	1	260	200	1	268	1	150	1	202	1	220	
228			200	1	268	1	158	1	186	1	242	

FIG. 8A

19/57

	ot201	PD	ca225		w3442		ca213		ga201		ga203	
1	178	1	160	1	14	1	112	1	150	1	188	1
1	170	1	160	1	14	1	124	1	150	1	208	1
1	170	1	160	1	18	1	124	1	154	1	208	1
1	170	1	184	1	22	1	112	1	158	1	212	1
0	186	0	174	0	18	0	124	1	150	0	212	0
0	170	0	160	0	14	0	124	1	146	0	188	0
1	170	1	178	1	18	1	128	1	146	1	192	1
1	178	1	176	1	22	1	112	1	154	1	216	1
1	170	1	172	1	10	1	126	0	146	1	188	1
1	186	1	172	1	16	1	124	0	158	1	188	1
1	170	1	160	1	10	1	124	1	150	1	212	1
1	170	1	180	1	14	1	124	1	150	1	196	1
1	178	1	176	1	10	1	126	1	150	1	188	1
1	170	1	176	1	18	1	124	1	150	1	188	1
1	178	1	174	1	14	1	124	1	154	0	192	1
1	170	1	160	1	14	1	124	1	150	0	188	1
1	178	1	170	1	16	1	130	1	154	0	216	1
1	170	1	160	1	14	1	128	1	150	0	192	1
1	186	1	172	0	10	1	112	1	150	1	212	1
1	170	1	174	0	20	1	124	1	158	1	188	1
1	178	1	172	1	10	1	124	1	154	1	188	1
1	170	1	174	1	16	1	126	1	146	1	188	1
1	178	1	176	1	10	1	126	1	154	1	208	1
1	178	1	172	1	10	1	126	1	142	1	212	1
1	178	0	174	0	18	0	124	1	154	1	216	1
1	170	0	172	0	10	0	126	1	158	1	188	1
0	186	0	174	0	18	1	124	1	158	0	212	0
0	178	0	172	0	18	1	124	1	146	0	208	0
1	178	0	172	0	26	0	124	1	158	1	216	1
1	170	0	168	0	10	0	124	1	158	1	188	1
0	178	0	172	0	10	0	124	1	150	0	188	1
0	170	0	174	0	16	0	124	1	142	0	188	1
1	170	1	174	0	18	1	128	1	150	1	192	1
1	178	1	172	0	18	1	124	1	158	1	208	1

FIG. 8B

20/57

Affected haplotypes

ca219		1105		ca209		ca202		1146		166d05		476	
241	1	85	1	173	1	192	1	272	1	312	1	271	1
233	1	99	1	181	1	196	1	270	1	304	1	271	1
241	1	85	1	173	1	182	1	274	0	312	1	273	1
245	1	103	1	177	1	194	1	270	0	308	1	267	1
241	1	85	1	173	1	198	0	274	1	308	0	275	0
241	1	85	1	173	1	194	0	274	1	304	0	271	0
241	1	87	1	173	1	182	1	272	1	300	1	271	1
235	1	101	1	181	1	196	1	274	1	312	1	271	1
235	1	85	1	173	1	182	1	274	1	312	1	271	1
243	1	85	1	173	1	192	1	274	1	316	1	267	1
245	1	103	1	177	1	194	0	274	0	312	1	271	1
235	1	91	1	181	1	182	0	270	0	316	1	271	1
241	1	85	1	173	1	182	1	274	1	312	1	271	1
241	1	103	1	177	1	196	1	274	1	312	1	271	1
241	1	85	1	173	1	196	0	270	1	300	1	271	0
235	1	85	1	181	1	190	0	274	1	312	1	267	0
235	1	81	1	173	1	182	1	274	1	324	1	271	0
223	1	83	1	173	1	192	1	274	1	300	1	267	0
245	1	103	1	177	1	196	1	274	1	312	1	271	1
241	1	85	1	173	1	182	1	270	1	312	1	265	1
241	1	105	0	173	1	196	1	270	1	304	1	267	1
241	1	101	0	173	1	196	1	270	1	308	1	271	1
241	1	87	0	173	1	192	1	274	1	312	1	271	1
241	1	85	0	173	1	196	1	274	1	304	1	267	1
245	1	97	1	177	1	194	1	274	1	312	0	271	1
235	1	99	1	181	1	198	1	274	1	300	0	271	1
241	0	95	0	181	0	198	0	274	1	320	0	273	0
235	0	85	0	173	0	196	0	274	1	308	0	271	0
235	1	101	0	181	0	196	1	272	1	312	1	271	1
235	1	85	0	173	0	200	1	274	1	308	1	271	1
241	0	85	0	173	1	200	0	274	0	312	0	271	0
243	0	101	0	173	1	196	0	270	0	304	0	267	0
241	1	85	1	173	1	182	1	274	1	316	1	271	0
241	1	99	1	173	1	200	1	274	1	300	1	269	0

FIG. 8C

21/57

Affected haplotypes

18p	PAN	MAN	ca212		1140		59		ca231		ta201
229	257	2	200	1	268	0	154	1	186	1	244
229			216	1	256	0	158	1	186	1	244
230	0	0	202	1	268	1	160	1	186	1	230
230			202	1	268	1	158	1	186	1	248
231	299	298	216	1	268	1	158	1	186	1	220
231			218	1	268	1	158	1	186	1	244
232	1	310	206	1	268	1	150	1	186	1	222
232			200	1	268	1	158	1	186	1	230
234	1	261	200	1	268	1	148	1	184	1	252
234			200	1	268	1	158	1	186	1	262
235	0	0	200	1	276	0	150	1	186	0	248
235			202	1	268	0	156	1	184	0	214
237	0	0	200	1	268	1	158	1	186	1	214
237			200	1	268	1	154	1	186	1	230
238	456	457	202	1	268	1	154	1	186	1	230
238			200	1	268	1	158	1	186	1	230
239	312	2	218	1	268	1	160	1	186	0	248
239			200	1	268	1	158	1	184	0	242
240	1	2	200	1	268	1	158	0	186	1	242
240			200	1	268	1	148	0	186	1	230
241	1	342	216	1	268	1	158	1	184	0	246
241			200	1	268	1	158	1	186	0	250
242	0	0	216	1	268	1	156	0	186	1	244
242			200	1	268	1	154	0	186	1	244
243	347	274	200	1	268	1	154	1	186	1	230
243			218	1	268	1	150	1	186	1	252
245	0	0	200	1	268	1	154	1	186	1	232
245			202	1	268	1	150	1	186	1	242
246	1	262	204	0	270	1	158	1	186	1	246
246			202	0	268	1	154	1	186	1	242
247	303	302	202	1	268	1	154	1	186	1	230
247			200	1	268	1	154	1	186	1	242
248	334	333	200	1	268	1	154	1	184	1	232
248			202	1	268	1	154	1	186	1	244

FIG. 8D

22/57

	at201	PD	ca225		w3442		ca213		ga201		ga203	
1	170	1	174	1	10	1	126	1	150	1	192	1
1	186	1	174	1	24	1	124	1	146	1	216	1
1	170	1	172	0	18	1	122	1	150	1	208	1
1	170	1	160	0	12	1	124	1	150	1	216	1
1	170	1	172	1	20	1	124	1	150	1	204	1
1	170	1	174	1	22	1	126	1	150	1	204	1
1	170	1	172	1	20	1	124	1	154	0	188	1
1	170	1	178	1	10	1	126	1	150	0	188	1
1	170	1	174	1	10	1	126	1	162	1	208	1
1	170	1	174	1	24	1	126	1	150	1	192	1
1	170	1	172	1	10	1	112	1	154	0	192	1
1	170	1	174	1	22	1	124	1	150	0	192	1
1	178	1	172	1	16	1	126	0	150	1	208	1
1	186	1	172	1	16	1	124	0	154	1	208	1
1	178	1	172	1	10	1	128	1	150	1	208	1
1	170	1	178	1	14	1	112	1	150	1	188	1
1	170	1	172	1	16	1	124	1	154	0	208	1
1	178	1	172	1	18	1	124	1	150	0	188	1
0	178	1	172	1	18	1	128	0	154	0	188	1
0	178	1	172	1	18	1	124	0	146	0	188	1
1	170	1	172	1	20	1	126	0	150	1	188	1
1	170	1	172	1	10	1	124	0	142	1	188	1
1	186	1	174	1	14	0	126	1	150	1	192	0
1	170	1	160	1	10	0	126	1	150	1	188	0
1	178	0	172	0	10	1	124	1	150	1	188	0
1	170	0	160	0	38	1	124	1	146	1	208	0
1	178	0	172	1	10	1	126	1	154	1	216	1
1	170	0	172	1	16	1	124	1	150	1	192	1
1	178	0	172	1	16	1	126	1	150	1	188	1
1	170	0	172	1	22	1	122	1	150	1	216	1
1	178	1	174	1	10	1	124	1	158	1	188	1
1	170	1	176	1	10	1	126	1	150	1	216	1
1	170	1	160	1	20	1	112	1	150	1		0
1	170	1	174	1	16	1	112	1	146	1		0

FIG. 8E

23/57

Affected haplotypes

ca219		1105		ca209		ca202		1146		166d05		476	
241	1	85	1	177	1	196	1	270	1	304	1	271	1
245	1	99	1	177	1	192	1	274	1	308	1	265	1
245	1	97	1	177	1	196	0	274	1	304	1	275	0
245	1	99	1	177	1	192	0	270	1	308	1	267	0
243	1	103	1	175	1	198	1	274	1	300	1	271	1
245	1	85	1	173	1	194	1	274	1	312	1	271	1
235	1	101	0	181	0	196	1	270	1	316	1	267	1
235	1	85	0	173	0	196	1	274	1	300	1	271	1
241	1	85	1	173	1	200	0	270	1	304	1	273	1
241	1	85	1	177	1	198	0	274	1	308	1	271	1
241	0	101	0	177	1	182	1	274	1	312	1	273	1
235	0	85	0	177	1	190	1	274	1	300	1	275	1
241	1	85	1	173	1	194	1	274	1	308	1	271	1
239	1	85	1	173	1	196	1	270	1	308	1	271	1
245	0	85	1	177	1	198	1	274	1	320	1	271	1
241	0	85	1	173	1	196	1	274	1	308	1	265	1
241	1	99	0	177	1	198	1	270	1	312	1	271	0
241	1	85	0	173	1	182	1	270	1	312	1	263	0
241	0	101	0	187	0	200	0	270	1	312	0	271	1
235	0	85	0	173	0	182	0	270	1	300	0	271	1
241	0	101	1	181	0	196	1	274	1	308	0	275	0
235	0	83	1	173	0	196	1	274	1	304	0	267	0
241	1	85	1	173	1	196	1	270	1	300	1	275	1
235	1	101	1	181	1	196	1	272	1	300	1	271	1
241	1	85	1	173	1	182	1	270	1	300	1	271	1
239	1	103	1	173	1	194	1	274	1	312	1	271	1
241	1	85	1	173	1	194	1	274	1	316	1	271	1
241	1	85	1	173	1	196	1	270	1	308	1	271	1
241	1	105	1	173	1	196	1	274	1	312	0	271	0
235	1	101	1	181	1	182	1	270	1	300	0	267	0
243	1	85	1	173	1	196	1	274	1	300	1	271	1
241	1	85	1	173	1	190	1	270	1	316	1	271	1
241	1	85	1	177	1	196	1	274	1	304	1	271	0
241	1	85	1	173	1	196	1	270	1	312	1	267	0

FIG. 8F

24/57

Affected haplotypes

18p	PAN	MAN	ca212		1140		59		ca231		ta201	
249	1	2	200	0	268	0	154	0	186	1	230	
249			216	0	256	0	148	0	186	1	246	
251	301	300	216	1	272	1	150	1	184	1	250	
251			216	1	268	1	158	1	186	1	244	
252	1	285	200	0	268	1	154	1	186	1	230	
252			204	0	268	1	158	1	186	1	246	
253	1	258	216	0	268	1	160	1	186	1	228	
253			200	0	268	1	154	1	186	1	230	
254	467	2	202	1	268	1	160	1	186	1	230	
254			200	1	268	1	154	1	186	1	230	
265	1	266	216	1	272	1	150	1	184	1	250	
265			202	1	268	1	154	1	186	1	230	
311	1	458	216	1	268	1	154	1	186	1	244	
311			200	1	268	1	162	1	186	1	242	
314	348	313	200	1	268	1	148	1	184	1	248	
314			216	1	268	1	162	1	184	1	250	
316	1	317	214	1	268	1	154	1	186	1	230	
316			200	1	268	1	154	1	186	1	242	
319	318	2	202	0	272	0	158	0	184	0	244	
319			200	0	256	0	154	0	186	0	244	
321	1	320	202	0	268	1	158	0		0		
321			200	0	268	1	154	0		0		
324	0	0	202	1	268	1	158	1	186	1	232	
324			216	1	268	1	150	1	196	1	220	
326	325	336	206	1	280	1	152	1	198	1	232	
326			202	1	268	1	154	1	186	1	232	
329	1	330	200	1	268	1	154	0	186	1	248	
329			206	1	268	1	148	0	186	1	234	
211	1	2	200	0	268	1	154	0	186	0	230	
211			204	0	268	1	148	0	198	0	252	
353	1	352	218	1	280	0	148	1	186	1	246	
353			200	1	268	0	148	1	186	1	246	
356	362	2	216	1	268	1	154	1	186	0	248	
356			204	1	268	1	164	1	190	0	232	

FIG. 8C

25/57

	at201	PD	ca225		w3442		ca213		ga201		ga203	
0	194	0	172	0	10	0	124	1	150	1	188	1
0	178	0	174	0	16	0	124	1	150	1	188	1
1	170	1	160	1	10	1	124	1	150	1	212	1
1	186	1	174	1	20	1	124	1	150	1	188	1
1	178	1	172	1	10	1	124	0	150	1	188	1
1	170	1	160	1	18	1	126	0	150	1	216	1
1	170	1	160	1	16	1	124	1	150	1	188	1
1	178	1	160	1	16	1	126	1	150	1	216	1
1	170	0	172	1	18	1	122	1	150	1	208	0
1	178	0	172	1	10	1	124	1	142	1	188	0
1	170	1	160	1	10	1	126	0	150	1	212	1
1	178	1	172	1	10	1	124	0	150	1	188	1
1	170	1	160	1	10	1	126	1	150	1	188	1
1	186	1	174	1	10	1	124	1	158	1	208	1
1	170	1	168	1	18	0	124	1	150	1	208	1
1	170	1	172	1	10	0	126	1	150	1	188	1
1	178	1	172	1	10	1	124	1	150	1	208	0
1	170	1	172	1	10	1	126	1	150	1	188	0
1	178	0	184	0	10	1	126	0	150	1	188	1
1	170	0	174	0	10	1	112	0	150	1	188	1
0	178	1	178	0	18	1	128	0		0		0
0	170	1	172	0	10	1	124	0		0		0
1	178	1	172	0	24	1	112	1	150	1	212	1
1	170	1	160	0	18	1	128	1	154	1	208	1
1	170	1	172	1	16	1	124	1	150	1	188	1
1	178	1	172	1	16	1	132	1	150	1	192	1
1	170	1	160	1	14	1	128	1	150	1	188	1
1	170	1	172	1	22	1	124	1	150	1	208	1
0	178	1	172	1	10	0	126	0	150	0	188	1
0	178	1	172	1	18	0	112	0	154	0	188	1
1	170	1	160	1	18	1	132	1	154	1	192	1
1	170	1	172	1	18	1	112	1	146	1	192	1
1	178	0	172	1	10	0	124	1	150	1	208	1
1	170	0	172	1	18	0	126	1	150	1	216	1

FIG. 8H

26/57

Affected haplotypes

ca219		1105		ca209		ca202		1146		166d05		476	
241	1	85	0	173	1	192	0	272	0	312	0	273	0
241	1	103	0	173	1	182	0	270	0	304	0	267	0
245	1	103	1	181	0	194	1	270	1	312	1	271	1
235	1	101	1	177	0	202	1	274	1	312	1	271	1
241	0	103	1	181	0	196	1	276	0	304	1	271	1
235	0	101	1	173	0	208	1	274	0	300	1	267	1
241	1	85	1	173	1	198	1	274	1	304	1	271	1
241	1	85	1	173	1	190	1	274	1	312	1	271	1
245	1	97	1	177	1	196	1	274	1	304	1	275	0
235	1	99	1	181	1	196	1	274	1	304	1	271	0
245	1	103	1	177	1	194	1	270	1	312	1	271	1
245	1	85	1	173	1	192	1	274	1	308	1	267	1
235	1	101	1	181	1	196	0	272	1	300	1	271	0
241	1	85	1	173	1	184	0	274	1	320	1	269	0
245	1	85	1	177	1	196	1	274	1	312	1	271	1
235	1	101	1	181	1	182	1	270	1	312	1	269	1
241	0	103	0	181	0	190	1	274	1	312	0	271	0
235	0	101	0	173	0	190	1	274	1	304	0	267	0
241	1	101	1	181	0	196	0	274	1	312	1	271	1
235	1	103	1	173	0	192	0	274	1	300	1	271	1
	0	101	1		0		0	270	1	304	1		0
	0	85	1		0		0	272	1	300	1		0
241	1	85	1	173	1	194	0	274	0	312	1	269	1
241	1	101	1	177	1	182	0	270	0	312	1	267	1
241	1	85	1	173	1	182	1	276	1	320	1	269	1
241	1	85	1	173	1	194	1	270	1	300	1	271	1
241	1	85	1	173	1	200	1	272	0	304	1	271	1
241	1	85	1	173	1	182	1	270	0	316	1	271	1
241	0	85	1	181	1	190	0	274	1	316	0	267	0
235	0	85	1	181	1	182	0	274	1	312	0	263	0
235	1	81	1	179	1	196	0	274	1	312	1	269	1
235	1	85	1	179	1	182	0	274	1	312	1	271	1
235	1	85	0	181	1	194	1	274	1	300	1	275	0
241	1	101	0	173	1	196	1	270	1	300	1	271	0

FIG. 8I

27/57

Affected haplotypes

18p	PAN	MAN	ca212		1140		59		ca231		ta201	
357	1	358	202	0	268	1	154	0	186	1	232	
357			214	0	278	1	158	0	186	1	248	
359	378	365	202	1	268	1	154	1	186	1	230	
359			202	1	272	1	158	1	184	1	244	
367	1	366	202	1	268	1	154	1	186	1	232	
367			202	1	268	1	154	1	186	1	242	
372	1	370	200	1	268	1	154	1	186	0		
372			216	1	268	1	148	1	184	0		
384	389	2	202	1	268	1	156	1	186	1	246	
384			202	1	268	1	154	1	186	1	250	
409	408	410	216	1	268	1	148	1	200	1	220	
409			202	1	268	1	154	1	186	1	230	
435	1	433	200	1	280	1	148	1	184	1	252	
435			202	1	268	1	156	1	194	1	220	
443	1	444	206	1	280	1	148	1	186	1	246	
443			202	1	256	1	154	1	186	1	230	
458	1	551	200	1	268	1	162	1	186	1	230	
458			200	1	268	1	154	1	186	1	234	
488	1	508	216	1	268	1	160	1	184	1	232	
488			216	1	268	1	160	1	184	1	232	
501	528	527	200	1	268	1	154	1	186	1	230	
501			206	1	268	1	154	1	186	1	244	
505	1	502	202	1	268	1	158	1	186	1	244	
505			200	1	268	1	158	1	186	1	244	
516	1	517	202	0	268	1	158	0		0		
516			200	0	268	1	148	0		0		
537	532	534	202	1	256	1	154	1	186	1	230	
537			216	1	268	1	154	1	184	1	230	
531	1	529	202	0	268	1	150	1	184	1	254	
531			200	0	268	1	154	1	186	1	244	
574	0	0	206	1	274	0	152	1	194	1	236	
574			200	1	268	0	148	1	184	1	252	
578	576	579	202	1	280	1	154	1	186	1	214	
578			202	1	268	1	154	1	186	1	230	

FIG. 8J

28/57

	at201	PD	ca225		w3442		ca213		ga201		ga203	
1	178	1	160	1	10	1	128	0	150	1	196	1
1	178	1	184	1	10	1	124	0	150	1	208	1
1	178	1	172	1	10	1	126	1	154	1	188	1
1	178	1	184	1	10	1	112	1	150	1	188	1
1	178	1	172	1	10	1	126	1	158	0	208	1
1	178	1	172	1	10	1	112	1	142	0	208	1
0		0	172	1	10	1	124	0	150	1		0
0		0	174	1	10	1	126	0	150	1		0
1	170	1	174	1	10	1	126	1	150	1	188	1
1	170	1	174	1	10	1	126	1	158	1	188	1
1	170	1	184	1	24	1	132	1	154	1	208	1
1	178	1	172	1	10	1	124	1	150	1	216	1
1	178	1	178	0	22	1	126	1	150	1	204	1
1	170	1	172	0	22	1	126	1	150	1	204	1
1	178	1	176	0	14	1	128	1	154	0	192	1
1	178	1	172	0	10	1	124	1	150	0	188	1
1	178	1	172	1	22	1	126	1	150	1	208	0
1	178	1	172	1	12	1	128	1	154	1	188	0
1	170	1	172	1	18	1	122	1	150	1	208	1
1	170	1	172	1	18	1	122	1	150	1	208	1
1	178	1	176	1	10	1	126	1	150	1	216	1
1	170	1	172	1	16	1	126	1	154	1	208	1
1	170	1	172	1	22	1	126	1	150	1	188	1
1	170	1	172	1	16	1	126	1	150	1	188	1
0		0		0	10	1	128	0		0	208	0
0		0		0	10	1	124	0		0	200	0
1	178	0	172	1	10	1	124	1	150	1	188	1
1	170	0	172	1	10	1	126	1	146	1	216	1
1	170	1	160	1	18	0	124	1	158	1	188	1
1	170	1	174	1	10	0	124	1	150	1	192	1
1	170	1	174	0	18	1	124	1	150	1	192	0
1	186	1	172	0	18	1	124	1	146	1	188	0
1	170	1	174	1	18	1	124	1	150	1	192	1
1	178	1	172	1	10	1	124	1	162	1	188	1

FIG. 8K

29/57

Affected haplotypes

ca219		1105		ca209		ca202		1146		166d05		476	
243	0	103	0	177	0	196	0	274	0	308	1	271	1
241	0	85	0	173	0	190	0	270	0	312	1	265	1
235	1	99	1	181	1	196	1	274	1	308	1	271	0
235	1	101	1	181	1	196	1	272	1	308	1	267	0
241	1	85	1	177	0	192	1	270	1	316	0	269	0
245	1	85	1	173	0	184	1	274	1	308	0	265	0
241	1	99	1	177	1		0	274	0	308	1	267	1
241	1	105	1	173	1		0	270	0	300	1	271	1
241	0	103	1	181	0	190	1	274	1	312	1	271	1
235	0	97	1	173	0	198	1	270	1	300	1	267	1
241	1	99	1	177	0	182	1	274	1	308	1	271	1
241	1	85	1	173	0	196	1	274	1	300	1	271	1
245	1	85	1	177	1	182	1	274	1	312	1	273	1
245	1	85	1	177	1	182	1	274	1	312	1	267	1
241	1	85	1	175	1	196	1	274	1	320	1	261	1
241	1	101	1	173	1	196	1	270	1	304	1	267	1
241	0	85	1	173	1	186	1	270	1	316	1	269	1
239	0	85	1	173	1	182	1	270	1	312	1	273	1
235	1	101	1	181	1	184	1	274	1	324	1	269	1
235	1	101	1	181	1	184	1	274	1	324	1	269	1
241	1	85	1	173	1	190	1	274	0	316	1	271	1
245	1	101	1	175	1	196	1	270	0	308	1	271	1
241	1	85	1	173	1	196	0	270	1	316	1	267	1
243	1	85	1	173	1	192	0	274	1	308	1	267	1
241	0	99	0	181	0	196	0	274	1	312	1	271	0
235	0	85	0	173	0	192	0	274	1	312	1	267	0
241	1	101	0	173	1	196	1	270	1	304	1	267	1
241	1	85	0	173	1	194	1	270	1	312	1	267	1
241	1	99	1	173	1	192	1	274	1	312	1	271	1
225	1	83	1	173	1	192	1	270	1	308	1	269	1
241	1	85	1	173	1	182	1	274	0	312	0	271	1
241	1	85	1	181	1	182	1	270	0	308	0	269	1
245	1	103	1	177	1	196	0	270	1	304	1	267	1
241	1	105	1	173	1	192	0	274	1	316	1	271	1

FIG. 8L

30/57

Affected haplotypes

18p	PAN	MAN	ca212		1140		59		ca231		ta201	
587	580	582	202	1	256	1	158	1	186	1	248	
587			202	1	268	1	154	1	186	1	244	
361	1	360	204	0	270	1	158	1	186	1	244	
361			202	0	276	1	148	1	186	1	236	
368	0	0	204	1	268	1	164	1	186	1	242	
368			202	1	256	1	154	1	186	1	230	
374	1	2	200	1	268	1	154	1	186	1	230	
374			200	1	268	1	154	1	186	1	230	
399	0	0	202	1	268	1	148	1	184	1		
399			204	1	272	1	158	1	186	1		
411	1	2	216	0	270	0	164	0	184	0	252	
411			202	0	268	0	154	0	186	0	230	
413	414	412	200	1	268	1	158	1	186	1	230	
413			202	1	280	1	148	1	186	1	244	
236	697	698	216	1	268	1	158	1	186	1	220	
236			216	1	268	1	158	1	186	1	220	
421	0	0	200	1	268	1	148	1	184	0	252	
421			202	1	268	1	152	1	186	0	242	
424	1	2	200	1	268	1	158	0	194	0	220	
424			200	1	268	1	154	0	186	0	232	
452	1	2	202	0	256	0	148	0	184	1	252	
452			200	0	268	0	154	0	184	1	250	
473	1	472	202	1	268	1	162	1	186	1	246	
473			218	1	268	1	148	1	186	1	244	
484	482	2	200	1	276	1	148	1	182	0	246	
484			206	1	256	1	154	1	186	0	244	
487	1	486	200	1	268	1	158	1	190	0	248	
487			202	1	278	1	148	1	186	0	246	
331	1	476	202	0	268	1	154	1	186	1	234	
331			200	0	268	1	154	1	186	1	230	
489	0	0	202	1	268	1	158	1	186	1	244	
489			200	1	268	1	148	1	202	1	220	
498	1	635	200	1	268	1	160	1	186	1	246	
498			200	1	268	1	164	1	186	1	246	

FIG. 8M

31/57

	at201	PD	ca225		w3442		ca213		ga201		ga203	
1	170	1	174	1	16	1	124	1	150	1	208	1
1	170	1	172	1	10	1	132	1	150	1	208	1
1	170	1	172	1	10	1	126	1	150	1	208	1
1	170	1	172	1	20	1	128	1	150	1	212	1
0	178	0	172	1	10	1	124	0	150	1	192	1
0	170	0	160	1	10	1	126	0	154	1	212	1
1	178	1	174	0	10	1	126	0	150	0	188	0
1	178	1	160	0	10	1	124	0	142	0	212	0
0	170	1	174	0	16	1	124	1	142	1	188	1
0	178	1	172	0	18	1	126	1	150	1	200	1
0	170	0	174	0	18	0	124	1	150	0	188	1
0	178	0	160	0	10	0	124	1	142	0	188	1
1	178	1	178	1	18	1	112	1	150	1	188	1
1	170	1	176	1	24	1	126	1	154	1	188	1
1	170	1	172	1	20	1	124	1	150	1	204	1
1	170	1	172	1	20	1	124	1	150	1	204	1
1	170	1	174	1	10	1	126	1	150	1	188	1
1	190	1	172	1	10	1	126	1	150	1	188	1
0	170	0	178	0	24	0	128	0	150	0	208	0
0	178	0	160	0	18	0	112	0	146	0	192	0
0	170	1	174	0	16	0	124	1	158	0	188	1
0	170	1	160	0	10	0	124	1	150	0	188	1
1	170	1	180	1	22	0	126	1	150	1	212	1
1	170	1	160	1	10	0	124	1	146	1	188	1
1	170	1	174	1	14	1	124	1	150	1	188	1
1	170	1	174	1	10	1	126	1	150	1	212	1
1	170	1	174	1	12	0	126	1	158	1	192	1
1	182	1	180	1	10	0	112	1	150	1	208	1
1	178	0	172	1	24	1	126	0	158	0	212	0
1	170	0	172	1	10	1	112	0	150	0	188	0
1	170	1	172	1	10	1	124	1	150	1	204	1
1	178	1	172	1	10	1	132	1	162	1	208	1
1	170	1	172	1	14	1	122	1	150	1	208	1
1	170	1	172	1	18	1	112	1	150	1	188	1

FIG. 8N

32/57

Affected haplotypes

ca219		1105		ca209		ca202		1146		166d05		476	
	0	85	1	173	1	190	1	274	1	312	1	271	1
	0	101	1	181	1	198	1	272	1	312	1	263	1
241	1	99	1	177	0	198	1	274	1	312	1	271	1
241	1	101	1	173	0	196	1	276	1	304	1	265	1
241	1	85	1	173	1	196	1	270	1	304	1	271	1
241	1	85	1	173	1	190	1	270	1	312	1	271	1
241	1	85	0	173	1	200	0	270	1	312	0	271	1
241	1	101	0	173	1	186	0	270	1	304	0	271	1
243	1	85	1	173	1	200	1	274	1	312	1	271	1
235	1	95	1	181	1	196	1	274	1	312	1	271	1
241	0	85	1	173	1	200	1	274	1	312	0	271	1
243	0	85	1	173	1	200	1	274	1	308	0	271	1
241	1	85	1	173	1	194	1	274	1	300	1	275	1
241	1	85	1	181	1	196	1	274	1	300	1	271	1
243	1	103	1	175	1	198	1	274	1	300	1	271	1
243	1	103	1	175	1	196	1	274	1	308	1	271	1
235	1	97	1	181	1	196	1	274	1	300	1	271	1
235	1	99	1	181	1	192	1	270	1	312	1	267	1
241	0	101	0	181	0	194	0	274	0	308	0	271	0
235	0	85	0	173	0	182	0	272	0	300	0	267	0
243	0	103	0	173	1	196	1	274	1	308	0	269	0
241	0	85	0	173	1	196	1	274	1	304	0	267	0
241	1	87	1	177	0	196	1	274	0	312	1	271	1
241	1	85	1	173	0	194	1	270	0	300	1	275	1
241	1	105	1	173	1	196	1	274	0	312	1	271	0
241	1	85	1	173	1	192	1	270	0	312	1	267	0
243	0	85	1	173	1	198	1	270	1	304	1	271	1
241	0	85	1	173	1	196	1	274	1	312	1	271	1
241	1	85	1	173	1	196	1	274	1	308	1	271	0
241	1	85	1	173	1	182	1	274	1	320	1	265	0
245	0	85	1	177	0	198	1	274	1	304	1	271	0
241	0	85	1	173	0	194	1	274	1	312	1	267	0
241	1	103	1	177	1	196	1	270	1	316	1	267	1
235	1	99	1	181	1	192	1	270	1	312	1	271	1

FIG. 80

33/57

Affected haplotypes

18p	PAN	MAN	ca212		1140		59		ca231		ta201	
566	0	0	216	1	268	1	148	1	202	1	220	
566			202	1	268	1	154	1	186	1	230	
514	1	2	202	0	268	1	154	1	186	0	230	
514			200	0	268	1	154	1	184	0	230	
536	1	633	202	1	270	0	148	1	184	1	254	
536			200	1	268	0	154	1	186	1	252	
605	1	2	216	0	268	1	158	0	198	0	244	
605			200	0	268	1	150	0	186	0	220	
540	539	562	200	1	268	1	154	1	186	1	230	
540			216	1	268	1	148	1	186	1	230	
684	1	730	202	0	268	1	158	1	186	1	232	
684			200	0	268	1	154	1	186	1	244	
608	1	2	206	0	268	1	156	0	192	0	244	
608			202	0	268	1	154	0	186	0	220	
637	1	638	216	1	268	1	162	0	186	1	250	
637			200	1	268	1	154	0	186	1	230	
649	647	646	200	1	268	1	154	1	186	1	230	
649			200	1	270	1	162	1	184	1	250	
653	1	652	200	1	280	0	160	0	184	1	230	
653			200	1	268	0	148	0	186	1	230	
491	1	2	204	0	268	1	158	0	194	0	256	
491			202	0	268	1	148	0	184	0	230	
493	1	2	202	0	282	0	158	0	186	1	242	
493			200	0	268	0	156	0	186	1	242	
506	1	2		0		0		0		0		
506				0		0		0		0		
661	660	662	200	1	278	1	156	1	198	1	220	
661			200	1	268	1	148	1	184	1	250	
667	666	668	202	1	268	1	154	1	186	1	214	
667			202	1	268	1	162	1	186	1	246	
669	670	671	202	1	268	1	162	1	186	1	258	
669			200	1	268	1	154	1	186	1	244	
676	1	678	202	0	268	1	158	1	190	1	244	
676			200	0	280	1	148	1	184	1	252	

FIG. 8P

34/57

	at201	PD	ca225		w3442		ca213		ga201		ga203	
0	178	1	174	1	10	1	124	1	150	1	212	1
0	178	1	172	1	10	1	128	1	150	1	208	1
1	178	1	172	0	10	1	124	1	154	0	192	1
1	178	1	168	0	10	1	124	1	146	0	192	1
1	170	1	168	0	16	1	132	0	162	1	212	0
1	178	1	172	0	24	1	124	0	154	1	188	0
0	170	0	172	1	16	0	124	1	158	0	200	0
0	178	0	172	1	10	0	124	1	150	0	188	0
1	178	1	172	1	10	1	124	1	150	1	216	1
1	194	1	172	1	22	1	112	1	154	1	212	1
1	178	1	160	1	24	1	112	1	150	1	212	1
1	170	1	160	1	10	1	126	1	150	1	188	1
0	170	1	178	0	22	0	126	1	150	1	204	0
0	170	1	174	0	10	0	126	1	150	1	188	0
1	182	1	172	1	10	1	124	1	142	0	208	1
1	178	1	172	1	10	1	124	1	150	0	212	1
1	178	1	172	1	10	1	124	1	150	1	188	1
1	170	1	180	1	10	1	112	1	154	1	188	1
1	178	1	184	1	20	1	128	1	154	1		0
1	178	1	168	1	22	1	112	1	150	1		0
0	178	0	180	0	22	0	124	1	158	0	204	0
0	170	0	174	0	10	0	124	1	154	0	188	0
1	170	1	174	0	16	0	124	1	158	0	212	0
1	170	1	172	0	14	0	124	1	150	0	204	0
0		0		0		0		0	150	1		0
0		0		0		0		0	150	1		0
1	170	1	174	1	20	1	126	1	154	1	204	1
1	186	1	174	1	18	1	120	1	150	1	188	1
1	170	1	160	1	22	1	124	1	146	1	212	1
1	178	1	172	1	18	1	112	1	158	1	188	1
1	186	1	174	1	18	1	126	1	150	1	188	1
1	170	1	160	1	10	1	126	1	150	1	188	1
1	178	1	172	1	16	1	126	1	158	1	188	1
1	178	1	172	1	22	1	126	1	150	1	216	1

FIG. 8Q

35/57

Affected haplotypes

ca219		1105		ca209		ca202		1146		166d05		476	
245	1	105	0	177	1	196	1	274	1	300	1	267	1
245	1	85	0	177	1	198	1	274	1	320	1	271	1
241	1	97	0	177	0	196	1	274	0	304	0	271	1
241	1	85	0	173	0	196	1	272	0	300	0	271	1
241	1	99	0	177	0	196	1	274	1	312	1	271	1
241	1	85	0	173	0	182	1	274	1	312	1	271	1
243	0	85	1	173	1	200	0	274	1	308	1	271	1
235	0	85	1	173	1	194	0	274	1	308	1	271	1
241	1	85	1	173	1	190	1	274	0	312	1	267	1
235	1	85	1	173	1	196	1	272	0	316	1	267	1
241	0	85	1	181	0	196	1	274	0	312	1	269	1
235	0	101	1	173	0	196	1	272	0	300	1	271	1
245	0	101	0	173	0	182	0	274	1	312	1	273	0
241	0	85	0	177	0	190	0	274	1	312	1	267	0
239	1	85	1	173	1	190	1	270	1	300	1	271	1
241	1	85	1	173	1	198	1	274	1	304	1	271	1
241	0	85	1	173	1	198	1	270	1	304	1	271	1
243	0	85	1	173	1	182	1	274	1	312	1	269	1
245	1	85	1	179	1	196	1	270	1	308	1	271	1
241	1	85	1	173	1	196	1	270	1	304	1	265	1
241	0	103	0	173	1	198	0	274	1	308	1	269	0
235	0	81	0	173	1	196	0	274	1	308	1	265	0
241	1	103	0	177	0	196	0	270	1	308	0	271	0
241	1	85	0	173	0	190	0	270	1	300	0	269	0
245	0		0		0		0		0		0		0
241	0		0		0		0		0		0		0
235	1	81	1	173	1	196	1	276	1	300	1	271	1
241	1	85	1	173	1	196	1	274	1	308	1	265	1
245	1	103	1	177	1	196	1	270	1	308	1	271	1
235	1	97	1	181	1	192	1	274	1	312	1	271	1
241	1	101	0	173	1	192	1	274	1	316	1	271	1
235	1	85	0	181	1	190	1	270	1	312	1	271	1
235	1	97	1	181	1	198	1	274	1	312	0	271	1
243	1	103	1	173	1	182	1	274	1	308	0	273	1

FIG. 8R

36/57

Affected haplotypes

18p	PAN	MAN	ca212	1140	59	ca231	ta201	at201	PD	ca225	w3442	ca213	ga201	ga203
681	1	2	202	0	162	0	186	0	186	0	174	0	126	0
681			200	0	154	0	186	0	178	0	172	0	124	0
351	354	2	202	1	154	1	186	1	178	1	172	1	126	1
351			216	1	156	1	186	1	186	1	174	1	124	1
355	1	2	216	0	158	0	190	0	170	0	172	0	126	0
355			204	0	152	0	186	0	178	0	172	1	124	0

FIG. 8S

37/57

Affected haplotypes

ca219		1105		ca209		ca202		1146		166d05		476	
241	1	103	0	177	0	190	0	270	0	304	0	271	1
241	1	85	0	173	0	196	0	274	0	312	0	271	1
241	1	101	1	173	1	192	1	274	1	312	1	271	0
245	1	105	1	177	1	194	1	274	1	320	1	267	0
241	1	103	0	177	0	196	1	274	0	316	0	267	1
241	1	85	0	173	0	196	1	270	0	304	0	267	1

*FIG. 8T*

38/57

nontransmitted chromosomes

ERSN	KID	sava5		ca211		ca212		1140		59		ca231
279	200	235	1	193	1	216	1	268	1	148	1	186
280	200	233	1	205	1	202	1	278	1	148	1	184
349	204	235	1	197	1	202	1	268	1	156	1	184
309	204	235	1	195	1	202	1	268	1	148	1	186
277	207	227	1	205	1	200	1	268	1	148	1	184
278	207	227	1	195	1	200	1	268	1	158	1	186
459	214	233	1	197	1	200	1	268	1	152	1	184
460	214	233	1	203	1	216	1	280	1	158	1	184
270	215	235	1	193	1	200	1	268	1	154	0	188
259	216	231	1	193	1	200	1	268	0	150	1	184
272	218	233	1	195	1	204	1	268	1	150	1	186
273	218	235	1	193	1	200	1	256	1	154	1	186
267	220	233	1	205	1	200	0	268	1	158	1	186
264	225	227	1	201	1	200	1	268	1	150	1	186
260	228	229	1	197	1	200	1	268	1	164	1	186
257	229	227	1	207	1	200	1	256	0	160	1	186
298	231	233	1	193	1	200	1	280	1	158	1	186
299	231	229	1	207	1	200	1	268	1	148	1	202
310	232	233	1	205	1	202	1	268	1	148	1	204
261	234	233	1	189	1	206	1	272	1	154	1	186
697	236	235	1	197	1	200	1	268	1	154	1	186
698	236	233	1	195	1	202	1	278	1	148	1	184
456	238	235	1	199	1	216	1	268	1	160	1	184
457	238	233	1	197	1	200	1	268	1	160	1	186
312	239	227	1	197	1	202	1	268	1	148	1	184
342	241	227	1	193	1	202	1	256	1	158	1	184
347	243	229	1		0	216	1	278	1	150	1	186
274	243	233	1	193	1	204	1	268	1	160	1	186
262	246	231	1	193	0	202	0	268	1	148	1	202
302	247	235	1	195	1	200	1	256	1	150	1	186
303	247	227	1	195	1	200	1	268	1	158	1	186
334	248	225	1	183	1	216	1	268	1	152	1	186
333	248	233	1	205	1	200	1	268	1	152	1	186
300	251	227	0	193	1	200	1	278	1	148	1	184

FIG. 9A

39/57

	ta201	at201	ca225	w3442	ca213	ga201	ga203	
1	246	1 194	1 172	1 16	1 124	1 150	1 188	1
1	252	1 170	1 172	1 20	1 124	1 150	1 192	1
1	252	1 170	1 172	1 20	1 120	1 150	1 216	1
1	244	1 170	1 172	1 16	1 124	1 142	1 192	1
1	252	1 186	1 174	1 18	1 124	1 146	1 212	1
1	230	1 178	1 168	1 20	1 124	1 150	1 200	1
1	248	1 186	1 174	1 10	1 124	1 142	1 208	1
1	248	1 170	1 184	1 16	1 124	1 146	1 216	1
1	246	1 170	1 160	1 24	1 124	1 150	0 188	1
0	254	1 186	1 172	1 10	1 124	1 150	0 188	1
1	248	1 178	1 172	0 22	1 126	1 146	1 188	1
1	230	1 178	1 172	0 10	1 124	1 142	1 188	1
1	244	1 170	1 160	1 14	1 124	1 158	1 188	1
1	242	1 170	0 168	0 10	0 126	1 150	1 192	1
1	250	1 178	1 172	0 14	1 112	1 154	1 188	1
1	246	1 170	1 172	1 14	1 122	1 150	1 208	1
1	232	1 178	1 172	1 12	1 112	1 154	1 188	1
1	220	1 170	1 160	1 14	1 112	1 158	1 208	1
1	220	1 170	1 160	1 24	1 112	1 150	0 188	1
1	250	1 178	1 174	1 18	1 126	1 158	1 188	1
1	230	1 186	1 174	1 10	1 112	1 150	1 208	1
1	252	1 170	1 172	1 20	1 120	1 150	1 216	1
1	248	1 170	1 172	1 16	1 124	1 150	1 208	1
1	230	1 170	1 172	1 18	1 122	1 150	1 208	1
0	246	1 170	1 178	1 24	1 112	1 150	0 208	1
0	250	1 170	1 174	1 10	1 124	0 146	1 188	1
1	244	1 170	0 160	0 10	1 112	1 150	1 188	0
1	244	1 170	0 160	0 14	1 124	1 162	1 188	0
1	230	1 170	0 172	1 22	1 124	1 150	1 208	1
1	242	1 170	1 172	1 10	1 126	1 150	1 192	1
1	230	1 178	1 168	1 14	1 128	1 150	1 188	1
1	230	1 178	1 176	1 10	1 126	1 150	1 188	0
1	230	1 178	1 172	1 10	1 124	1 142	1 188	0
1	252	1 170	1 172	1 18	1 120	1 150	1 216	1

FIG. 9B

40/57

nontransmitted chromosomes

ca219		1105		ca209		ca202		1146		166d05		476	
241	1	103	1	173	1	186	1	274	1	316	1	269	1
241	1	85	1	173	1	182	1	270	1	316	1	263	1
243	1	85	1	177	1	192	1	270	0	312	1	265	1
241	1	85	1	173	1	192	1	270	0	312	1	267	1
241	1	85	1	173	1	198	1	274	1	308	1	271	1
245	1	101	1	175	1	196	1	274	1	316	1	267	1
245	1	101	1	177	1	190	1	274	1	312	1	267	1
241	1	85	1	173	1	202	1	270	1	312	1	269	1
235	1	95	1	181	1	190	0	274	1	308	1	267	0
241	1	85	1	173	1	196	1	274	1	304	1	267	0
235	1	103	1	181	1	196	1	274	1	312	1	265	1
235	1	99	1	181	1	196	1	274	1	304	1	271	1
241	1	85	0	173	1	192	1	270	1	312	1	271	1
241	1	85	0	173	0	196	1	270	1	304	1	271	1
235	1	93	1	181	1	196	1	274	1	308	1	269	0
241	1	103	1	177	1	196	1	270	1	316	1	267	1
235	1	97	1	181	1	198	1	274	1	300	1	271	1
245	1	85	1	177	1	192	1	270	1	300	1	271	1
241	1	85	0	173	0	196	1	274	1	308	1	271	1
235	1	95	1	181	1	198	0	274	1	300	1	267	1
241	1	85	1	173	1	196	1	274	1	300	1	267	1
243	1	85	1	177	1	192	1	270	1	312	1	265	1
241	0	99	1	177	1	198	1	270	1	312	1	263	1
241	0	97	1	177	1	196	1	274	1	304	1	275	1
245	1	85	0	177	1	196	1	272	1	308	1	263	0
235	0	99	1	173	0	196	1		0	304	0		0
235	1	101	1	181	1	194	1	274	1	308	1	267	1
241	1	85	1	177	1	196	1	274	1	304	1	271	1
245	1	85	1	177	1	198	1	270	1	300	0	267	0
241	1	85	1	173	1	196	1	270	1	304	1	271	1
239	1	85	1	181	1	196	1	276	1	300	1	267	1
241	1	85	1	181	1	194	1	274	1	324	1	267	0
241	1	99	1	181	1	196	1	274	1	304	1	267	0
243	1	85	1	177	0	192	1	270	1	312	1	265	1

FIG. 9C

41/57

nontransmitted chromosomes

ERSN	KID	sava5		ca211		ca212		1140		59		ca231
301	251	227	0	205	1	200	1	276	1	148	1	184
285	252	231	1	193	1	200	0	268	1	148	1	184
258	253	229	1	193	1	200	0	268	1	148	1	186
467	254	229	1	197	1	216	1	280	1	160	1	184
266	265	227	0	195	1	202	1	268	1	160	1	186
485	311	227	1	205	1	200	1	268	1	158	1	184
313	314	227	1	195	1	202	1	268	1	162	1	186
348	314	227	1	195	1	200	1	268	1	148	1	184
317	316	227	1	201	1	202	1	268	1	152	1	186
318	319	227	0		0		0	256	0	154	0	
320	321	237	1	201	0	200	0	268	1	154	0	186
336	326	227	1	193	1	202	1	268	1	154	1	186
325	326	227	1	201	1	202	1	276	1	148	1	186
330	329	233	1	197	1	202	1	268	1	148	0	184
476	331	229	0	199	1	200	0	276	1	154	1	
354	351	233	1	201	0	200	1	268	1	162	1	186
352	353	225	0	207	1	200	1	268	0	154	1	194
362	356	231	1	195	1	202	1	268	1	154	1	186
358	357	235	1	205	1	202	0	256	1	154	0	186
365	359	233	1	205	1	200	1	268	1	162	1	186
378	359	231	1	201	1	202	1	268	1	162	1	186
360	361	227	0	195	1	202	0	268	1	162	1	186
366	367	227	1	193	1	202	1	268	1	154	1	186
370	372	227	0	201	1	202	1	268	1	150	1	184
389	384	231	1	203	1	204	1	272	1	158	1	186
408	409	229	1	205	1	216	1	276	1	154	1	186
410	409	229	1	197	1	204	1	272	1	158	1	186
414	413	227	1	195	1	200	1	268	1	158	1	186
412	413	235	1	193	1	200	1	256	1	156	1	186
433	435	227	1	195	1	202	1	266	1	154	1	186
444	443	235	1	205	1	200	1	268	1	158	1	186
551	458	235	1	201	1	206	1	268	1	148	1	184
472	473	233	1	193	1	200	1	268	1	156	1	186
482	484	233	0	197	1	200	1	268	1	158	1	182

FIG. 9D

42/57

	ta201	at201	ca225	w3442	ca213	ga201	ga203	
1	252	1 170	1 172	1 24	1 124	1 150	1 188	1
1	252	1 170	1 174	1 16	1 124	0 150	1 192	1
1	230	1 194	1 172	1 22	1 112	1 154	1 208	1
1	250	1 170	0 172	1 22	1 126	1 154	1 188	0
1	260	1 178	1 174	1 16	1 124	0 158	1 208	1
1	230	1 178	1 184	1 20	1 128	1 154	1 212	1
1		0 170	1 172	1 10	0 124	1 150	1 212	1
1	248	1 170	1 172	1 10	0 128	1 150	1 208	1
1	244	1 170	1 174	1 14	1 112	1 154	1 188	0
0		0	0	0 16	1	0	0	0
0	220	1 170	1 172	0 20	1 124	0 146	0 192	0
1	244	1 170	1 160	1 18	1 124	1 154	1 208	1
1	244	1 170	1 176	1 20	1 126	1 150	1 192	1
1	256	1 178	1 172	1 16	1 124	1 162	1 208	1
0	244	1 170	0 160	1 10	1 112	0 150	0 188	0
1	248	1 178	1 160	1 22	1 132	1 150	1 188	1
1	220	1 170	1 178	1 18	1 128	1 146	1 192	1
0	230	1 170	0 172	1 10	0 128	1 150	1 208	1
1	230	1 178	1 172	1 10	1 124	0 154	1 216	1
1	248	1 178	1 160	1 22	1 132	1 150	1 188	1
1	230	1 186	1 174	1 18	1 126	1 150	1 188	1
1	250	1 170	1 172	1 18	1 124	1 150	1 212	1
1	230	1 178	1 160	1 10	1 124	1 142	0 188	1
0	244	0 170	1 174	1 14	1 124	0 150	1 188	1
1	244	1 178	1 172	1 18	1 126	1 150	1 200	1
1	244	1 178	1 184	1 28	1 112	1 154	1 196	1
1	244	1 178	1 172	1 18	1 126	1 150	1 188	1
1	242	1 178	0 174	1 18	1 120	1 150	1 188	1
1	246	1 170	1 172	1 10	1 124	1 150	1 212	1
1	242	1 170	1 172	0 16	1 112	1 150	1 204	1
1	232	1 178	1	0 24	1 112	1 150	0 188	1
1	248	1 170	1 174	1 14	1 124	1 158	1 188	0
1	248	1 178	1 184	1 10	0 112	1 146	1 188	1
0	248	1 170	1 174	1 16	1 124	1 150	1 188	1

FIG. 9E

43/57

nontransmitted chromosomes

ca219		1105		ca209		ca202		1146		166d05		476	
235	1	101	1	177	0	200	1	272	1	316	1	267	1
235	0	85	1	173	0	192	1	274	0	308	1	267	1
235	1	101	1	181	1	196	1	274	1	308	1	265	1
245	1	103	1	175	1	198	1	274	1	300	1	271	0
235	1	101	1	181	1	202	1	274	1	316	1	265	1
245	1	85	1	179	1	184	0	270	1	308	1	269	0
241	1	85	1	173	1	192	1	270	1	312	1	269	1
241	1	85	1	173	1	198	1	270	1	308	1	271	1
235	0	101	0	173	0	190	1	274	1	304	0	267	0
245	1	85	1		0		0	274	1	320	1	269	1
	0	103	1	173	1	182	1	274	1	312	1	271	0
241	1	85	1	173	1	182	1	270	1	312	1	273	1
241	1	85	1	177	1	200	1	274	1	308	1	263	1
235	1	85	1	173	1	196	1	270	0	316	1	265	1
241	1	85	1	173	1	182	1	270	1	300	1	265	0
241	1	85	1	173	1	182	1	270	1	308	1	267	0
241	1	87	1	173	1	182	0	272	1	300	1	271	1
245	1	85	0	177	1	198	1	274	1	300	1	271	0
241	0	85	0	173	0	190	0	270	0	312	1	273	1
241	1	85	1	173	1	182	1	270	1	312	1	267	0
241	1	85	1	177	1	192	1	270	1	312	1	267	0
241	1	85	1	173	0	192	1	270	1	308	1	269	1
243	1	85	1	173	0	200	1	274	1	308	0	265	0
243	1	85	1	173	1	190	0	270	0	316	1	273	1
235	0	95	1	173	0	196	1	274	1	312	1	271	1
243	1	85	1	173	0	198	1	270	1	300	1	271	1
241	1	85	1	173	0	196	1	274	1	316	1	267	1
241	1	85	1	173	1	194	1	270	1	316	1	265	1
241	1	85	1	173	1	200	1	274	1	316	1	271	1
241	1	85	1	173	1	194	1	270	1	300	1	271	1
235	1	105	1	181	1	200	1	272	1	316	1	267	1
239	0	101	1	173	1	196	1	274	1	300	1	271	1
241	1	85	1	173	0	192	1	270	0	316	1	265	1
241	1	83	1	173	1	196	1	270	0	304	1	267	0

FIG. 9F

44/57

nontransmitted chromosomes

ERSN	KID	sava5		ca211		ca212		1140		59		ca231
486	487	227	1	201	1	202	1	256	1	154	1	186
508	488	233	1	205	1	202	1	268	1	148	1	184
635	498	227	1	193	1	202	1	268	1	148	1	184
527	501	229	1	183	1	216	1	280	1	158	1	186
528	501	225	1	183	1	216	1	268	1	152	1	186
502	505	235	1	205	1	200	1	268	1	148	1	184
517	516		0		0		0		0		0	
529	531	233	1	205	1	200	0	268	1	158	1	186
633	536	229	0	201	1	200	1	268	0	154	1	186
532	537	227	1	201	1	200	1	268	1	150	1	186
534	537	235	1	205	1	200	1	268	1	158	1	186
562	540	229	1	195	1	202	1	268	1	160	1	184
539	540	229	1	207	1	200	1	268	1	154	1	194
576	578	235	1	199	1	200	1	256	1	158	1	186
579	578	233	1	199	1	200	1	278	1	148	1	186
582	587	227	1	201	1	202	1	268	1	148	1	202
580	587	229	1		0	200	1	268	1	154	1	186
638	637	237	1	203	1	206	1	268	1	154	0	186
647	649	229	1	195	1	202	1	268	1	154	1	186
646	649	231	1	201	1	206	1	268	1	154	1	186
652	653	235	1	201	1	206	1	268	0	154	0	186
662	661	235	1	209	1	202	1	280	1	154	1	186
660	661	233	1	183	1	216	1	268	1	158	1	186
666	667	235	1	203	1	202	1	268	1	158	1	186
668	667	237	1	209	1	202	1	268	1	150	1	186
670	669	235	1	205	1	200	1	268	1	148	1	184
671	669	227	1	195	1	200	1	268	1	158	1	186
678	676	223	1	201	1	200	0	278	1	156	1	200
730	684	229	1	195	1	200	0	268	1	148	1	198

FIG. 9C

45/57

		ta201		at201		ca225		w3442		ca213		ga201		ga203	
	0	230	1	178	1	172	1	10	0	124	1	150	1	188	1
	1	220	1	170	1	160	1	24	1	112	1	150	1	188	1
	1	254	1	170	1	174	1	16	1	124	1	142	1	188	1
	1	230	1	170	1	172	1	22	1	126	1	146	1	212	1
	1	242	1	170	1	176	1	10	1	126	1	154	1	208	1
	1		0	170	1	174	1	10	1	126	1	150	1	188	1
	0		0		0		0		0		0		0		0
	1	242	1	170	1	180	1	10	0	128	1	150	1	204	1
	1	230	1	178	1	168	0	10	1	124	0	150	1		0
	1	242	1	170	0	172	1	10	1	126	1	150	1	192	1
	1	232	1	170	0	160	1	24	1	112	1	150	1	208	1
	1	250	1	170	1	160	1	18	1	124	1	150	1	212	1
	1	220	1	170	1	178	1	18	1	128	1	150	1	192	1
	1	246	1	170	1	174	1	10	1	124	1	158	1	188	1
	1	246	1	170	1	184	1	16	1	124	1	150	1	208	1
	1	220	1	178	1	184	1	10	1	128	1	150	1	212	1
	1	244	1	170	1	160	1	10	1	126	1	150	1	200	1
	1	228	1	170	1	160	1	22	1	126	1	142	0	212	1
	1	232	1	178	1	160	1	10	1	124	1	150	1	216	1
	1	230	1	178	1	160	1	10	1	124	1	154	1	188	1
	1	230	1	178	1	172	1	16	1	126	1	150	1	188	0
	1	242	1	178	1	172	1	22	1	126	1	150	1	188	1
	1		0	170	1	160	1	14	1	122	1	150	1	192	1
	1	246	1	170	1	174	1	10	1	126	1	150	1	192	1
	1	252	1	178	1	172	1	16	1	128	1	150	1	196	1
	1	254	1	170	1	174	1	10	1	126	1	154	1	192	1
	1	230	1	178	1	168	1	16	1	128	1	154	1	188	1
	1	252	1	174	1	174	1	10	1	124	1	150	1	208	1
	1	220	1	170	1	174	1	20	1	126	1	150	1	196	1

FIG. 9H

46/57

nontransmitted chromosomes

ca219		1105		ca209		ca202		1146		166d05		476	
241	0	103	1	173	1	192	1	274	1	312	1	267	1
243	1	85	1	173	1	196	1	274	1	308	1	273	1
243	1	85	1	173	1	200	1	274	1	312	1	271	1
241	1	85	1	173	1	182	1	270	0	320	1	267	1
241	1	87	1	173	1	198	1	270	0	312	1	267	1
235	1	97	1	181	1	192	0	274	1	300	1	271	1
	0		0		0		0	272	1		0		0
235	1	81	1	173	1	182	1	278	1	320	1	261	1
241	1	85	0	173	0	200	1	270	1	304	1	271	1
241	1	85	0	173	1	196	1	270	1	304	1	271	1
235	1	85	0	181	1	194	1	274	1	308	1	267	1
239	1	85	1	173	1	194	1	272	0	316	1	271	1
241	1	85	1	173	1	182	1	272	0	300	1	271	1
241	1	105	1	173	1	192	0	274	1	312	1	267	1
241	1	87	1	173	1	192	0	272	1	304	1	275	1
	0	103	1	173	1	194	1	270	1	316	1	271	1
	0	101	1	173	1	196	1	272	1	308	1	271	1
241	1	87	1	173	1	182	1	274	1	320	1	269	1
241	0	85	1	173	1	194	1	270	1	312	1	267	1
241	0	85	1	173	1	196	1	274	1	300	1	271	1
235	1	99	1	181	1	192	1	274	1	312	1	267	1
235	1	101	1	181	1	196	1	272	1	300	1	271	1
235	1	85	1	179	1	196	1	274	1	312	1	271	1
241	1	85	1	173	1	192	1	270	1	312	1	271	1
241	1	87	1	173	1	182	1	270	1	316	1	273	1
235	1		0	181	1	196	1	274	1	300	1	271	1
239	1	85	0	181	1	196	1	276	1	300	1	267	1
241	1	83	1	177	1	182	1	276	1	308	0	269	1
235	0	93	1	173	0	202	1	272	0	300	1	273	1
					79								

FIG. 9I

47/57

controls

cont	sava5	ca211		ca212		1140		59		ca231		ta201
98	miss	193	1	200	1		0	156	1	186	1	230
98	17	193	1	216	1		0	148	1	186	1	244
99		193	1	206	1	268	1	150	1	184	1	252
99		195	1	200	1	268	1	154	1	184	1	220
101		189	1	206	1	272	1	154	1	186	1	260
101		203	1	200	1	268	1	150	1	186	1	244
102		195	1	202	1	268	1	150	1	202	1	220
102		205	1	200	1	268	1	162	1	186	1	248
104		195	1	200	1	268	1	154	1	186	1	244
104		203	1	216	1	268	1	156	1	186	1	244
105		193	1	202	1	268	1	156	1	186	1	244
105		201	1	216	1	268	1	148	1	186	1	246
107			0	206	1	268	1	154	1	186	1	246
107			0	202	1	274	1	150	1	184	1	246
108		201	0	200	1	268	1	162	1	186	1	230
108		195	0	202	1	280	1	154	1	186	1	242
110		199	1	218	1	268	1	160	1	184	1	248
110		205	1	200	1	268	1	148	1	184	1	254
111		193	1	202	1	268	1	154	1	186	1	232
111		191	1	202	1	268	1	150	1	184	1	252
114		207	1	202	1	268	1	150	1	200	1	220
114		195	1	200	1	278	1	154	1	186	1	252
113		191	1	202	1	276	1	150	1	184	1	250
113		207	1	216	1	268	1	150	1	186	1	244
116		193	1	202	1	268	1	154	1	186	1	230
116		195	1	202	1	268	1	154	1	186	1	248
117		201	1	200	1	268	1	154	1	186	1	232
117		195	1	202	1	268	1	160	1	186	1	256
119		193	1	200	1	270	1	162	1	186	1	244
119		193	1	206	1	268	1	154	1	186	1	230
120		193	1	216	1	276	1	158	1	186	1	242
120		203	1	204	1	272	1	158	1	186	1	244
122		183	1	200	1	268	1	154	1	186	1	242
122		195	1	218	1	268	1	156	1	186	1	232

FIG. 10A

48/57

	at201	ca225	w3442	ca213	ga201	ga203	
1	178	176	10	126	150	208	1
1	178	172	18	124	150	208	1
1	178	172	20	124	142	204	1
1	170	170	10	128	150	188	1
1	178	174	20	126	158	216	1
1	170	160	14	122	150	188	1
1	178	172	24	124	150	212	1
1	178	160	22	132	150	188	1
1	170	160	10	126	150	188	1
1	186	174	14	126	150	192	1
1	170	172	10	126	150	188	1
1	194	172	16	124	150	188	1
1	170	176	22	0	154	188	1
1	170	174	16	0	150	216	1
1	178	172	22	126	150	188	1
1	178	172	22	126	150	192	1
1	170	172	16	0 124	150	208	1
1	170	174	10	0 126	150	188	1
1	178	160	0	124	150	188	1
1	170	160	0	128	150	188	1
1	170	174	0 24	126	150	212	1
1	178	172	0 18	124	150	192	1
1	170	174	0 22	124	146	216	1
1	170	172	0 16	124	150	192	1
1	178	172	10	124	0 150	188	1
1	170	172	10	126	0 150	212	1
1	178	172	10	124	0 142	212	1
1	178	174	16	126	0 158	212	1
1	170	172	18	0 124	150	216	1
1	178	172	10	0 126	150	188	1
1	178	174	18	0 112	154	192	1
1	178	172	10	0 126	150	200	1
1	178	160	16	124	150	204	1
1	178	160	26	124	150	188	1

FIG. 10B

49/57

controls

ca219		1105	18SCA20	SC	KID
241	1	missing	173	1	100
241	1		177	1	100
235	1		173	1	100
245	1		175	1	100
235	1		173	1	103
235	1		181	1	103
241	1		173	1	103
241	1		173	1	103
235	1		181	1	106
241	1		173	1	106
241	1		173	1	106
241	1		173	1	106
241	1		173	1	109
241	1		173	1	109
241	1		173	1	109
245	1		175	1	109
235	1			0	112
235	1			0	112
243	1		181	0	112
235	1		173	0	112
241	1		173	1	115
241	1		173	1	115
241	1		173	1	115
241	1		173	1	115
241	1		173	1	118
241	1		173	1	118
241	1		173	1	118
241	1		177	1	118
241	1		173	1	121
241	1		173	1	121
241	1		173	1	121
235	1		181	1	121
	0		173	1	124
	0		173	1	124

FIG. 10C

50/57

controls

cont	sava5	ca211		ca212		1140		59		ca231		ta201
123		193	1	200	1	268	1	150	1	184	1	252
123		195	1	216	1	268	1	154	1	184	1	232
125		203	1	200	1	268	1	148	1	184	1	252
125		205	1	202	1	268	1	148	1	188	1	250
126		205	1	200	1	268	1	148	1	186	1	248
126		195	1	204	1	268	1	150	1	186	1	246
128		193	1	200	1	256	1	158	1	186	1	
128		191	1	200	1	268	1	160	1	184	1	
129		193	1	206	1	256	1	154	1	186	1	244
129		195	1	216	1	268	1	150	1	184	1	250
131		201	0	200	1	268	1	154	1	186	0	252
131		197	0	200	1	268	1	150	1	184	0	244
132		205	0	200	1	268	1	148	1	186	0	252
132		203	0	200	1	268	1	158	1	184	0	248
134		193	1	216	1	268	1	148	1	186	1	220
134		205	1	202	1	266	1	160	1	186	1	230
135		193	1	202	1	268	1	154	1	186	1	244
135		205	1	202	1	268	1	154	1	184	1	230
138		193	1	202	1	268	1	154	1	186	1	230
138		207	1	200	1	280	1	148	1	184	1	252
137		193	1	206	1	268	1	154	1	186	1	230
137		201	1	216	1	270	1	148	1	184	1	256
144			0	200	1	256	1	154	1	186	1	
144			0	206	1	268	1	154	1	186	1	
68		195	1	202	1	268	1	164	1	186	1	
68		193	1	202	1	268	1	160	1	186	1	
69		195	1	218	1	268	1	148	1	186	1	246
69		201	1	216	1	268	1	158	1	186	1	230
72		193	1	200	1	268	1	148	1	184	1	
72		193	1	206	1	256	1	156	1	186	1	
71		193	1	216	1	268	1	146	1	192	1	248
71		193	1	206	1	256	1	156	1	186	1	232
74		195	1	218	1	268	1	148	1	186	1	246
74		205	1	200	1	268	1	158	1	186	1	222

FIG. 10D

51/57

		at201		ca225		w3442		ca213		ga201		ga203	
	1	170	1	160	1	10	1	126	1	154	0	188	1
	1	170	1	160	1	20	1	112	1	150	0	192	1
	1		0	174	1	18	1	124	1		0	212	1
	1		0	172	1	16	1	124	1		0	192	1
	1	170	1	160	1	14	1	128	0	150	1	188	1
	1	178	1	172	1	22	1	126	0	150	1	208	1
	0	170	1	174	1	14	1	112	1	158	1	188	1
	0	170	1	172	1	18	1	122	1	150	1	208	1
	1	170	1	174	1	10	1	112	1	158	1	188	1
	1	170	1	172	1	14	1	126	1	150	1	192	1
	0	186	0	176	0	18	1	126	1	150	1	188	1
	0	170	0	172	0	10	1	126	1	150	1	188	1
	0	186	0	174	0	18	1	124	1	150	1	212	1
	0	170	0	172	0	18	1	124	1	158	1	208	1
	1	170	1	174	1	14	1	124	1	150	1	208	1
	1	194	1	172	1	22	1	112	1	154	1	208	1
	1	170	1	160	1	18	1	124	1	154	1	208	1
	1	178	1	184	1	20	1	128	1	154	1	208	1
	1	178	1	172	0	10	1	124	1		0	216	1
	1	178	1	174	0	20	1	126	1		0	216	1
	1	178	1	172	0	10	1	126	1	150	1	192	1
	1	186	1	174	0	10	1	126	1	150	1	212	1
	0		0	174	1	10	1	126	1	150	1	208	1
	0		0	176	1	22	1	124	1	150	1	188	1
	0		0	172	1	22	1	126	1	150	1	208	1
	0		0	172	1	18	1	122	1	150	1	208	1
	1		0	160	1	10	1	124	1	146	1	208	1
	1		0	172	1	20	1	124	1	150	1	204	1
	0	170	1	174	1	16	1	124	1	150	1	188	1
	0	170	1	172	1	10	1	124	1	150	1	192	1
	1	170	1	174	1	16	1	124	1	154	1	196	1
	1	170	1	174	1	10	1	126	1	150	1	212	1
	1	170	1	160	1	10	1	124	1	154	1	216	1
	1	170	1	160	1	24	1	112	1	154	1	188	1

FIG. 10E

52/57

controls

co219		1105	18SCA20	SC	KID
241	1		173	1	124
241	1		177	1	124
241	1		173	1	127
241	1		173	1	127
243	1		173	1	127
239	1		173	1	127
235	1		181	1	130
235	1		181	1	130
241	1		173	1	130
241	1		173	1	130
243	0		181	0	133
235	0		173	0	133
245	0		181	0	133
235	0		173	0	133
243	1		173	1	136
235	1		181	1	136
241	1		173	1	136
241	1		173	1	136
241	1			0	139
243	1			0	139
235	1		181	0	139
241	1		177	0	139
241	1		173	1	145
241	1		173	1	145
241	1		173	1	70
245	1		177	1	70
241	1		173	1	70
243	1		175	1	70
243	1		173	1	73
241	1		173	1	73
235	1		181	1	73
241	1		173	1	73
241	1		173	1	76
235	1		181	1	76

FIG. 10F

53/57

controls

cont	sava5	ca211		ca212		1140		59		ca231		ta201
75		217	1	216	1	264	1	150	1	186	1	250
75		205	1	204	1	268	1	154	1	186	1	244
78		201	1	216	1	268	1	148	1	186	1	
78		201	1	202	1	268	1	162	1	186	1	
77		201	1	206	1	268	1	158	1	184	1	246
77		195	1	202	1	268	1	152	1	186	1	232
80		193	1	202	0	268	1		0	186	1	250
80		195	1	200	0	268	1		0	186	1	244
81		193	1	202	0	268	1	156	1	186	1	246
81		193	1	200	0	268	1	148	1	184	1	258
84		193	1	202	1	268	1	154	1	186	1	246
84		207	1	202	1	268	1	164	1	186	1	244
83		209	1	200	1	270	1	148	1	184	1	230
83		207	1	200	1	268	1	158	1	186	1	248
86		195	1	202	1	268	1	158	1	186	1	244
86		205	1	202	1	278	1	148	1	184	1	260
87		197	1	200	1	268	1	158	1	186	1	230
87		193	1	200	1	268	1	154	1	190	1	242
90		205	1	200	1	268	1	158	1	186	1	250
90		193	1	200	1	268	1	154	1	186	1	246
89		207	1	202	1	270	1	168	1	186	1	232
89		193	1	202	1	268	1	154	1	190	1	252
92		193	1	200	1	268	1	148	1	184	0	244
92		193	1	202	1	256	1	154	1	186	0	230
93		203	1	216	1	268	1	156	1	186	0	248
93		205	1	200	1	268	1	148	1	184	0	230
95		197	1	216	1	268	1	158	1	186	1	252
95		205	1	202	1	268	1	150	1	184	1	230
96		209	1	200	1	278	1	162	1	186	1	256
96		205	1	200	1	268	1	148	1	186	1	230
140			0		0	270	1		0		0	244
140			0		0	278	1		0		0	254
141		201	0	200	1	272	1		0		0	244
141		193	0	200	1	270	1		0		0	254

FIG. 10G

54/57

		at201		ca225		w3442		ca213		ga201		ga203	
	1	170	1	180	1	12	1	124	1	150	1	192	1
	1	170	1	172	1	16	1	124	1	146	1	192	1
	0	174	1	172	1		0	124	1	150	1	192	0
	0	170	1	174	1		0	126	1	150	1	188	0
	1	170	1	160	1	22	1	124	1	150	1	192	0
	1	178	1	174	1	20	1	122	1	146	1	188	0
	1	178	1	160	1	10	1	124	1	150	1	208	1
	1	178	1	172	1	28	1	124	1	150	1	208	1
	1	194	1	172	1	10	1	126	1	150	1	188	1
	1	186	1	174	1	10	1	124	1	150	1	208	1
	1	170	1	172	1	14	1	126	1	158	1	188	1
	1	170	1	178	1	10	1	124	1	150	1	188	1
	1	178	1	172	1	26	1	124	1	150	1	208	1
	1	170	1	174	1	10	1	112	1	146	1	192	1
	1	170	1	160	1	14	1	124	1	158	1	208	1
	1	170	1	172	1	20	1	120	1	150	1	188	1
	1	178	1	172	1	10	1	124	1	158	1	188	1
	1	170	1	172	1	16	1	126	1	154	1	188	1
	1	170	1	172	1	18	1	124	1	158	1	208	1
	1	186	1	172	1	10	1	124	1	150	1	188	1
	1	178	0	176	1	22	1	126	1	154	1	212	1
	1	170	0	172	1	16	1	126	1	150	1	188	1
	1	170	0	174	1	10	1	124	1	150	1	208	1
	1	178	0	172	1	10	1	124	1	154	1	188	1
	1	170	1	174	1	14	1	126	1	150	1	204	1
	1	178	1	174	1	10	1	126	1	150	1	188	1
	1	178	1	174	1	20	1	126	1	150	1	192	0
	1	178	1	160	1	10	1	126	1	150	1	188	0
	1	170	1	160	1	14	1	128	1	150	1	192	0
	1	178	1	160	1	14	1	128	1	150	1	188	0
	1		0		0	10	1		0	150	1	188	1
	1	186	1		0	10	1		0	158	1	188	1
	1	170	1	172	1	10	1		0		0	216	1
	1	170	1	160	1	10	1		0		0	212	1

FIG. 10H

55/57

controls

ca219		1105	18SCA20	SC	KID
241	1		175	1	76
235	1		177	1	76
241	1		173	1	79
241	1		177	1	79
241	1		173	1	79
241	1		181	1	79
241	1			0	82
241	1			0	82
241	1		173	1	82
241	1		173	1	82
241	1		173	1	85
241	1		173	1	85
241	1		173	1	85
241	1		173	1	85
	0		173	1	88
	0		173	1	88
235	1		173	1	88
235	1		181	1	88
245	1		173	1	91
241	1		173	1	91
241	1		181	1	91
241	1		173	1	91
241	1		173	1	94
241	1		173	1	94
241	1		177	1	94
235	1		181	1	94
241	1		173	1	97
245	1		173	1	97
221	1		173	1	97
241	1		173	1	97
	0			0	142
	0			0	142
241	1		173	1	142
241	1		173	1	142

FIG. 10I

controls

ca219		1105	18SCA20	SC	KID
241	1		177	1	145
235	1		181	1	145

FIG. 10K

56/57

controls

cont	save5	ca211	ca212	1140	59	ca231	td201	at201	ca225	w3442	ca213	ga201	ga203
143		193	1 200	1 278	1 148	1 184	1 252	1 170	1 184	1 18	1 124	1 150	1 192
143		195	1 200	1 268	1 158	1 186	1 248	1	178	1 18	1 124	1 146	1 188

FIG. 10J

57/57

## AHR RESULTS IN DISEASE CHROMOSOMES

	SAVA5	CA211	CA212	18S1140	18S59	TA201	CA231	AT201	CA225	W3442
SAVA5	0.04 2%				0.07 4%					
CA211	2-3									
CA212										
18S1140					0.70 12%		1.92 30%		1.28 20%	
18S59	2-6			2-4		2.11 10%	2.45 18%	0.53 10%	1.16 12%	4.18 17%
TA201					4-3		0.68 3%	1.34 10%	0.02 2%	0.68 8%
CA231				2-2	4-2	2-4			0.89 16%	
AT201					4-2	3-2			0.18 4%	0.03 4%
CA225				2-3	4-3	3-3	2-3	1-1		
W3442					4-1	3-1	2-1	2-1		

FIG. 11

## INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/14892

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12Q 1/68; C12P 19/34; C07H 21/02, 21/04

US CL : 435/6; 91.2; 536/23.1, 24.3, 24.33

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6; 91.2; 536/23.1, 24.3, 24.33

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

Please See Extra Sheet.

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Coon et al. Analysis of Chromosome 18 DNA Markers in Multiplex Pedigrees with Manic Depression. Biological Psychiatry. 15 April 1996. Volume 39, pages 689-696, see entire document.	1-16
A	Stine et al. Evidence for Linkage of Bipolar Disorder to Chromosome 18 with a Parent-of-Origin Effect. American Journal of Human Genetics. December 1995. Volume 57, pages 1384-1394, see entire document.	1-16

☐ Further documents are listed in the continuation of Box C.☐ See patent family annex.

* Special categories of cited documents:	*T* later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
*A* document defining the general state of the art which is not considered to be of particular relevance	*X* document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
*E* earlier document published on or after the international filing date	*Y* document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
*L* document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*Z* document member of the same patent family
*O* document referring to an oral disclosure, use, exhibition or other means	
*P* document published prior to the international filing date but later than the priority date claimed	

Date of the actual completion of the international search

28 SEPTEMBER 1997

Date of mailing of the international search report

17 DEC 1997

Name and mailing address of the ISA/US  
Commissioner of Patents and Trademarks  
Box PCT  
Washington, D.C. 20231

Facsimile No. (703) 305-3230

Authorized officer

DIANNE REES

Telephone No. (703) 308-0196

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US97/14892

## B. FIELDS SEARCHED

Electronic data bases consulted (Name of data base and where practicable terms used):

APS, BIOSIS, BIOTECHABS, BIOTECHDS, CABA, CAPLUS, DISSABS, DRUGU, EMBASE, EUROPATFULL, EUROPEX, MEDLINE, SCISEARCH, WPIDS, TOXLINE, TOXLIT, USPATFULL

search terms: bipolar, manic depression, mood disorder, shizoaffective disorder, chromosome 18, 18p, short arm, and markers, D18S1140, ga203, SAVA5, W3422, at201, ta201,